再版前言

本书第一版受到出乎意料的欢迎：各兄弟院校广泛采用；发行20余万册；获第一届国家级优秀教材奖．这应归功于时代的需要，也说明作者试图写出一本既有前苏联教本的系统严谨，又有美国书籍的生动活泼，体现我国教学经验，理论与应用兼备的教材的目标得到一定程度的实现．

广大师生的热情反映和意见，写作时就存在的不少遗憾，陆续发现的一些不足甚至错误，18年来在教学和科研中的些许心得，这些成了第二版修改的依据．

本书的结构受到一致的肯定，这次未加更改，只是重写了过于薄弱的数字特征部分，因而增加一节，并恢复了初版时因篇幅考虑而删去的全书小结．

增写了不少典型应用事例，补写上几个影响学科初期发展的名例，改写了一些平淡无味的实例，这些使本书的理论与应用更为均衡，更为尊重历史事实，也更准确地反映概率论这一学科目前在整个自然科学和社会科学中的重要地位．相信也增添了本书的趣味性．

一种分布就是一个数学模型．初版重视分布及它们之间联系的传统得到发扬．正式引入负二项分布与埃尔兰分布，使两个等待分布序列的对比更为明显．改动了一些例题与习题．这样一来，本书把分布按重要性分成三类：第一类三大分布有专门的节加以介绍；第二类包括十几种重要分布，在正文中指明其背景、性质以及它与其他分布的关系；第三类则多数在习题中出现．最后在附录一中汇总．关于多元分布，新版也较前重视．

几个定理的证明被局部更动，多数为改善，少数为改正．增

添了唯一的一个定理（移植自浅野、江島两位先生与作者合著的一本日文书），用极其简练的办法证明了重要的多元中心极限定理.

习题是本书的重要组成部分. 这次略作调整，基本题约占四分之三，与正文紧密配合，标有星号的题目对正文作了补充，双星号的是难题. 为适应不同需要，配题较多，教学中选用约半数即可.

以上是关系全局的一些较大变动，其他增删所在多是. 定理、公式、图表的编号作了统一编排. 关键词附上英文，重要数学家的名字也都标出原文，译名有些更动. 希望通过这些修改能使原书的质量有所提高. 预料书中仍会存在不少缺点与错误，欢迎广大师生批评、指正.

修改后的第二版篇幅略有增加，如为学时所限，建议舍去相对独立的熵与信息一节和相当专深的最后两节.

在我的概率论生涯中，得到郑绍濂、吴立德、陶宗英、汪嘉冈、何声武、卞国瑞、徐家鹄等师友的许多帮助，高尚华为本书的前后两版付出大量心血，缪铨生教授主持了第二版的审稿，在此一并致谢.

李贤平

1996 年中秋

目　录

第一章　事件与概率

§1. 随机现象与统计规律性

一、随机现象

概率论（probability theory）是研究随机现象的数量规律的数学分支．为了说明什么是随机现象，让我们先来看一个例子．设某车间有 200 台车床，由于经常需要检修、测量、调换刀具、变换位置等种种原因，因此，即使在生产期间，各台车床还是时常需要停车．若每台车床有百分之六十的时间在开动，而每台车床开动时需要耗电 1 千瓦，问应供给这个车间多少电力才能保证此车间正常生产？

类似的例子在许多实际问题中出现，解决这类问题当然具有重要意义．

显然，若供给这个车间 200 千瓦的电能则此车间便能正常生产．但这样做不合算，因为每台车床的开工率只有 60%，也就是说，平均起来这个车间中同时在工作的车床只有 120 台，供给它们 200 千瓦的电能太多．供给 120 千瓦的电能行吗？这又太少些，因为有时工作的车床数会超过 120 台，若只供给 120 千瓦，则这时会出现因电力不足而使车床无法正常运转，那么到底供给多少才能既保证生产正常进行而又节约电力呢？

用概率论的方法能给这个问题以完满的解决，现在我们仅把结果列出，详细的解法以后再讲．计算表明，只要供给这个车间 141 千瓦的电就够了，虽然在这时也可能碰到因电力不足而不能正常生产的情况．但这种机会很少，它小于 0.1%，即在 8 小时工

作中一般只有半分钟会碰到这种情况，这显然影响不大．但节约出来的 59 千瓦电能却能用于许多别的用途．

解决这个问题的关键在于要计算出某时刻同时工作着的车床数，但是由于某台车床在某时刻是否开工很难预先确定，这受到许多偶然因素的影响，即工作着的车床数是一个受许多偶然因素影响的量．

原来，在自然界和人类社会中存在着两类不同的现象．

当我们多次观察自然现象和社会现象后，会发现许多事情在一定的条件下必然会发生．例如在没有外力作用的条件下，作等速直线运动的物体必然继续作等速直线运动；又如在海边生活时，水加热到 100℃时必然会沸腾等等．这种在一定条件下，必然会发生的事情称为**必然事件**．反之，那种在一定条件下，必然不会发生的事情就称为**不可能事件**．例如在不受外力作用的条件下，作等速直线运动的物体改变其等速直线运动状态是不可能的．

从所举例子中看出，必然事件和不可能事件，虽然形式相反，但是两者的实质是相同的．必然事件的反面就是不可能事件，而不可能事件的反面就是必然事件．

所有这种现象我们称之为**决定性现象**，它广泛地存在于自然现象和社会现象中．

但是在自然现象和社会现象中也还广泛存在着与决定性现象有着本质区别的另一类现象，上述车间供电问题就是一例．

类似的例子还可以举出很多，例如用同一仪器多次测量同一物体的重量，所得结果彼此总是略有差异，这是由于诸如测量仪器受大气影响，观察者生理上或心理上的变化等等偶然因素引起的．同样地，同一门炮向同一目标发射多发同种炮弹，弹落点也不一样，因为炮弹制造时种种偶然因素对炮弹质量有影响，此外，炮筒位置的误差，天气条件的微小变化等等都影响弹落点．再如从某生产线上用同一种工艺生产出来的灯泡的寿命也有差异等等．总之，所举这些现象的一个共同的特点是：在基本条件不变

的情况下，一系列试验或观察会得到不同的结果. 换句话说，就个别的试验或观察而言，它会时而出现这种结果，时而出现那种结果，呈现出一种偶然性. 这种现象称为**随机现象**. 对于随机现象通常关心的是在试验或观察中某个结果是否出现，这些结果称为**随机事件**，简称**事件**（event). 例如过马路交叉口时可能遇上各种颜色的交通指挥灯，这是一个随机现象，而“遇到红灯”则是一个随机事件. 以后我们用 A，B，C 等大写拉丁字母表示随机事件.

二、频率稳定性

正如恩格斯所指出的，在表面上是偶然性在起作用的地方，这种偶然性始终是受内部的隐蔽着的规律支配的，而问题只是在于发现这些规律.

人们经过长期的实践发现，虽然个别随机事件在某次试验或观察中可以出现也可以不出现，但在大量试验中它却呈现出明显的规律性——频率稳定性.

对于随机事件 A，若在 N 次试验中出现了 n 次，则称

$$F_N(A)=\frac{n}{N}$$

为随机事件 A 在 N 次试验中出现的**频率**.

下面是关于频率稳定性的几个有名例子. 援引这类例子是因为它们不但具有一定的权威性，而且都是可以反复验证的.

在掷一枚硬币时，既可能出现正面，也可能出现反面，预先作出确定的判断是不可能的，但是假如硬币均匀，直观上出现正面与出现反面的机会应该相等，即在大量试验中出现正面的频率，应接近于50%，为了验证这点，历史上曾有不少人做过这个试验，其结果如下页所示①.

又如，在英语中某些字母出现的频率远远高于另外一些字母.

---

① 引自格涅坚科. 概率论教程. 高等教育出版社，第44页.

实　验　者	掷硬币次数	出现正面次数	频　　率
蒲　　丰	4040	2048	0.5069
皮　尔　逊	12000	6019	0.5016
皮　尔　逊	24000	12012	0.5005

在进行了更深入的研究之后，人们还发现各个字母被使用的频率相当稳定. 例如，下面就是英文字母使用频率的一份统计表①. 其他各种文字也都有着类似的规律.

字母	空格	E	T	O	A	N	I	R	S
频率	0.2	0.105	0.072	0.0654	0.063	0.059	0.055	0.054	0.052
字母	H	D	L	C	F	U	M	P	Y
频率	0.047	0.035	0.029	0.023	0.0225	0.0225	0.021	0.0175	0.012
字母	W	G	B	V	K	X	J	Q	Z
频率	0.012	0.011	0.0105	0.008	0.003	0.002	0.001	0.001	0.001

近年来对汉语的统计研究有了很大的发展. 关于汉字的使用频率已有初步统计资料，对汉语常用词也作了一些统计研究. 特别是结合汉字输入方案等的研制，正在对汉字的结构作深入的统计分析. 这些研究对实现汉字信息处理自动化无疑具有重要的意义.

另一个验证频率稳定性的著名试验是由英国生物统计学家高尔顿(Galton)设计的. 它的试验模型如图 1 所示.

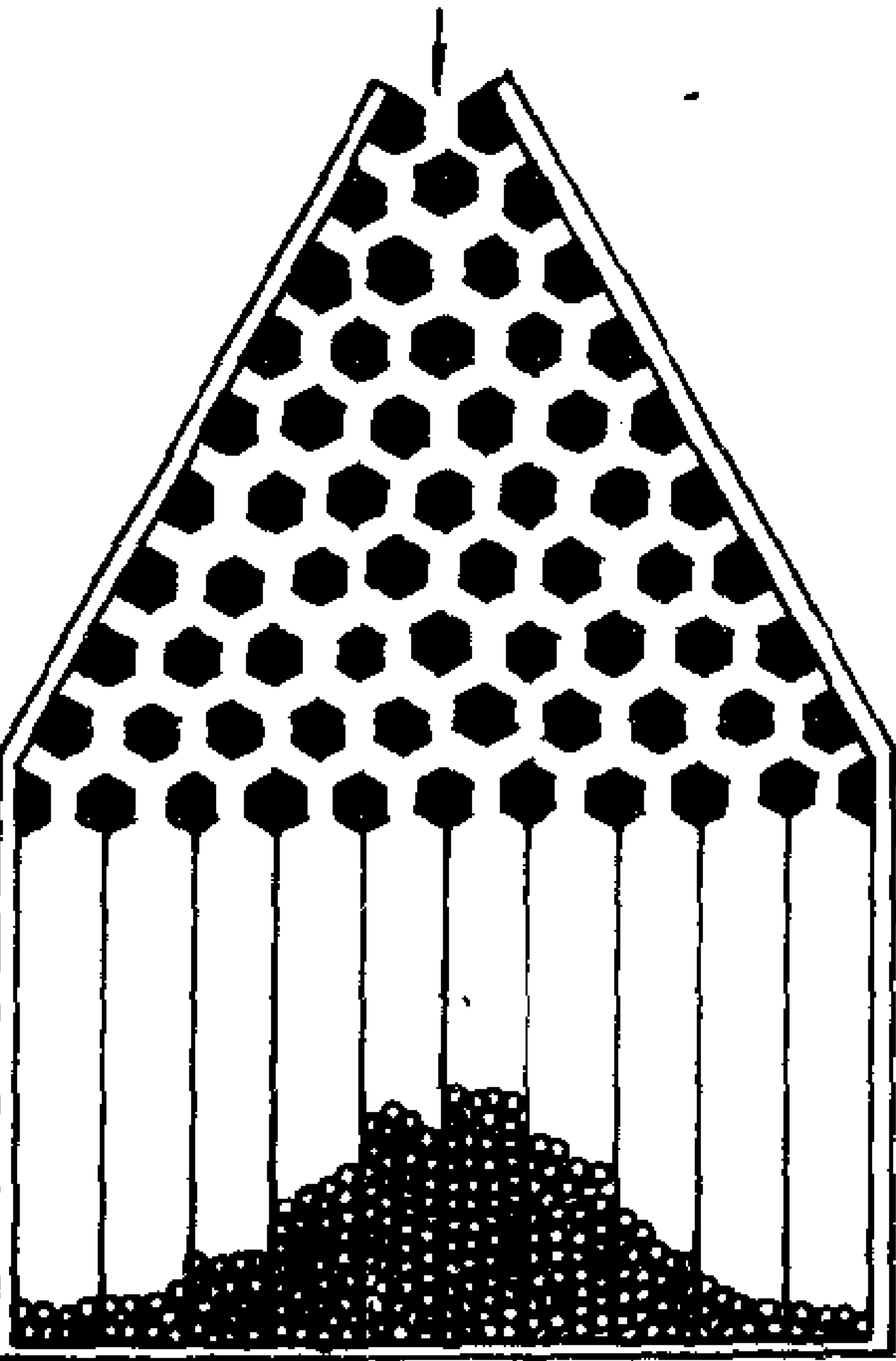


图 1　高尔顿板

自上端放入一小球，任其自由下落，在下落过程中当小球碰到钉子时，从左边落下与从右边落下的机会相等. 碰到下一排钉子时又是如此. 最后落入底板中的某一格子. 因此，任意放入一球，则此球落入

① 引自 L. Brillouin. Science and Information Theory. New York, 1956.

哪一个格子，预先难以确定. 但是实验证明，如放入大量小球，则其最后所呈现的曲线，几乎总是一样的. 也就是说，小球落入各个格子的频率十分稳定. 这个试验模型称为**高尔顿板**. 试验中呈现出来的规律性，在学习第五章极限定理之后，就会有更深刻的理解.

同样，如果多次测量同一物体，其结果虽略有差异，但当测量次数增加时，就会越来越清楚地呈现出一些规律性：测量值的平均值在某固定常数附近波动，诸测量值在此常数两旁的分布呈现某种对称性. 又如在射击的例子中，当射击次数不多时，炮弹的弹落点似乎是前后左右杂乱无章，看不出什么明显的规律；但当射击次数增加时，弹落点的分布就会呈现出一定的规律性：如弹落点关于目标的分布略呈对称性，偏离目标远的弹落点比偏离目标近的弹落点少等等. 其他如灯泡寿命等，在进行多次观察或试验后，也都可以发现类似的规律性.

日常生活中也不乏有趣的例子，例如衣服和用具总在同样部位以相似的方式破损，下雨时地面各处总是差不多同时淋湿等等. 读者只要多注意观察，就不难发现许多关于频率稳定性的有说服力的实例.

上述种种事实表明，随机现象有其偶然性的一面，也有其必然性的一面. 这种必然性表现为大量试验中随机事件出现的频率的稳定性，即一个随机事件出现的频率常在某个固定的常数附近摆动，这种规律性我们称之为**统计规律性**. 频率的稳定性说明随机事件发生的可能性大小是随机事件本身固有的、不随人们意志而改变的一种客观属性，因此可以对它进行度量.

对于一个随机事件 A，用一个数 $P(A)$ 来表示该事件发生的可能性大小，这个数 $P(A)$ 就称为随机事件 A 的**概率**(probability). 因此概率度量了随机事件发生的可能性的大小.

对于随机现象，只讨论它可能出现什么结果，价值不大，而指出各种结果出现的可能性的大小则具有很大意义. 有了概率的

概念就使我们能对随机现象进行定量研究，由此建立了一个新的数学分支——概率论.

三、频率与概率

既然概率 $P(A)$ 度量了随机事件 A 发生的可能性大小，可以预料，在 N 次重复试验中，若 $P(A)$ 较大，则频率 $F_N(A)=\frac{n}{N}$ 也较大. 反之若 $P(A)$ 很小，则 $F_N(A)$ 也很小，而且概率 $P(A)$ 应与频率有许多相似的性质. 以下我们先对频率的性质进行一番考察.

首先，频率具有非负性

$$F_N(A)\geqslant 0 \tag{1.1.1}$$

其次，对于必然发生的事件，在 N 次试验中应出现 N 次. 若以 Ω 记必然事件，则应有

$$F_N(\Omega)=1 \tag{1.1.2}$$

还有，若 A 及 B 是两个不会同时发生的随机事件，以 $A+B$ 表示 A 或 B 至少出现其一这个事件，则应有

$$F_N(A+B)=F_N(A)+F_N(B) \tag{1.1.3}$$

这个性质称为**频率的可加性**.

当然还可以列出频率的许多性质，但上述三个性质是最基本的. 例如，“不可能事件在 N 次试验中出现的频率为 0”，“任何随机事件在 N 次试验中出现的频率不大于 1”，“对于有限个两两不会同时发生的随机事件也有频率可加性”，这些性质都可以由 (1.1.1)，(1.1.2) 及 (1.1.3) 推出.

最后，根据上述频率稳定性的讨论似乎可以提出这样的猜想，即当 N 足够大时 $F_N(A)$ 与 $P(A)$ 应充分接近. 这一点有很大的启发性，在历史上它一直是概率论研究的一个重大课题. 以后我们将会看到，在很一般的条件下，这个结论的确成立，但同时还须对问题的提法进一步明确化.

频率与概率的上述关系有时还提供了求某事件概率的一种手段，即当 N 足够大时，用它的频率来作为概率的近似值. 以后我们将会看到，这种做法大有好处.

四、概率论简史

概率论是一门研究随机现象数量规律的学科，一般把1654年作为概率论诞生的一年. 这年中，法国数学家巴斯卡与费马就机会博弈中的一些问题作了通信讨论；后来惠更斯也加入研究. 在这些研究中建立了概率论的一些基本概念，如事件、概率、数学期望等.

其后，随着生产实践的发展，特别是在射击、保险、测量等工作中提出的一些概率问题，促使人们在概率论的极限定理等方面进行深入研究，起初主要对伯努利试验概型进行研究，而后则推广到更为一般的场合. 这个时期先后对概率论作出重要贡献的数学家有伯努利、拉普拉斯、泊松和高斯.

极限定理的研究在18世纪和19世纪整整200年中成了概率论研究的中心课题. 在本世纪初，由于新的更有力的数学方法的引入，这些古典问题得到了较好解决，这当中俄罗斯彼得堡学派起了主导作用.

虽然概率论历史悠久，但是它的严格的数学基础的建立以及理论研究和实际应用的极大发展却主要是本世纪的事情.

由于物理学（如统计物理）、生物学以及工程技术（如自动电话、无线电技术）发展的推动，概率论得到了飞速的发展. 理论课题不断扩大与深入，概率论的思想渗入各个学科成了近代科学发展的明显的特征之一.

由于各个数学分支的发展与互相渗透，概率论的严格的数学基础被建立起来，古典问题得到了解决，新的概念和工具不断出现，概率论成了数学的一个活跃分支.

概率论大大地扩大了它的应用范围. 特别在最近几十年中，概

率论的方法被引入各个工程技术学科和社会学科. 目前，概率论在近代物理，无线电与自动控制，工厂产品的质量管理，医药和农业试验，金融保险业等等方面都找到了重要应用，这些实际需要也有力地推动了概率论的新发展，有些还形成了边缘学科（如信息论、排队论、可靠性理论等）.

在这个时期内，由于生物学和农业试验的推动，数理统计学（mathematical statistics）也获得了很大发展，它以概率论为理论基础，又为概率论应用提供了有力的工具，两者互相推动，迅速发展. 而概率论本身的研究则转入以随机过程为中心课题，取得了许多理论上和应用上都有重要价值的结果.

§2. 样本空间与事件

一、样本空间

从本节开始，我们将逐步引进概率论的基本概念. 样本空间与事件是最基本的两个概念.

对随机现象的研究必然要联系到对客观的事物进行“调查”、“观察”或“实验”，以后我们统称之为**试验**（trial），并假定这种“试验”可以在相同条件下重复进行.

我们感兴趣的是试验的结果. 例如掷一次硬币，我们关心的是出现正面或出现反面，这是两个可能出现的结果. 假如我们考察的是掷两次硬币的试验，则可能出现的结果有（正，正），（正，反），（反，正），（反，反）四种；如果掷三次硬币，则结果还要复杂，但还是可以把它们描述出来. 总之，为了研究随机试验，首先需要知道这个试验可能出现的结果. 这些结果称为**样本点**，一般用 ω 表示. 样本点全体构成**样本空间**（sample space），用 Ω 表示. 在具体问题中，给定样本空间是描述随机现象的第一步.

下面举一些例子.

［例 1］　在研究英文字母使用情况时，把样本空间选为 $\Omega=$

{空格，A，B，…，Z} 是适宜的，这个样本空间只有有限个样本点，是比较简单的样本空间.

[例 2] 观察一小时中落在地球上某一区域的粒子数，可能的结果一定是非负整数，而且很难指定一个数作为它的上界，这样，可以把样本空间取为 $\Omega=\{0, 1, 2, \cdots\}$. 这个样本空间含有无穷多个样本点，但这些样本点可以依照某种次序排列出来，以后我们将称它的点数为可列个.

[例 3] 讨论某地区的气温时，我们自然把样本空间取为 $\Omega=(-\infty, \infty)$，或 $\Omega=[a, b]$. 这个样本空间包含有无穷多个样本点，它们充满一个区间，不是一个可列集.

[例 4] 考察地震震源时，可以把样本点取为 (x, y, z)，其中 x 表示震源的经度，y 表示纬度，z 表示深度. 这时，样本空间是三维空间中某一区域.

从以上例子可以看出，随着问题不同，样本空间可以相当简单，也可以相当复杂.

在今后讨论中，经常把样本空间认为是预先给定的. 当然对于一个实际问题或一个随机现象，如何用一个恰当的样本空间来描述它也很值得研究. 但是在概率论的研究中，一般都假定样本空间是给定的. 这是必要的抽象，这种抽象使我们能更好地把握住随机现象的本质，而且得到的结果能广泛地应用. 事实上，一个样本空间可以概括各种实际内容很不相同的问题：例如只包含两个样本点的样本空间既能作为掷硬币出现正、反面的模型，也能用于产品检验中出现“好品”及“废品”，又能用于气象中“下雨”与“不下雨”，以及公用事业排队现象中“有人排队”与“无人排队”等等. 尽管问题的实际内容如此不同，但有时却能归结为相同的概率模型. 我们后面常以摸球等作为例子也是由于这个原因，它能使问题的本质更为突出.

二、事件

有了样本空间的概念，就可以定义事件．我们还是从考察一个例子开始．

［例 5］　口袋中装有 4 只白球和 2 只黑球，我们考虑依次从中摸出两球所可能出现的事件．若对球进行编号，4 只白球分别编为 1，2，3，4 号，2 只黑球编为 5，6 号．如果用数对 (i, j) 表示第一次摸得 i 号球，第二次摸得 j 号球，则可能出现的结果是

$$\begin{array}{l}(1,2),(1,3),(1,4),(1,5),(1,6)\\(2,1),(2,3),(2,4),(2,5),(2,6)\\(3,1),(3,2),(3,4),(3,5),(3,6)\\(4,1),(4,2),(4,3),(4,5),(4,6)\\(5,1),(5,2),(5,3),(5,4),(5,6)\\(6,1),(6,2),(6,3),(6,4),(6,5)\end{array} \quad (*)$$

把这 30 个结果作为样本点，则构成了样本空间．在这个问题中，这些样本点是我们感兴趣的事件；但是我们也可以研究下面另外一些事件：

A：第一次摸出黑球；

B：第二次摸出黑球；

C：第一次及第二次都摸出黑球．

后面这些事件与前面那些事件的不同处在于这些事件是可以分解的，例如为了 A 出现必须而且只须下列样本点之一出现：

$$\begin{array}{l}(5,1),(5,2),(5,3),(5,4),(5,6)\\(6,1),(6,2),(6,3),(6,4),(6,5)\end{array}$$

前面的 30 个事件由单个样本点构成；后面这三个事件，每一个事件都是由若干个样本点构成的，总之，它们都是样本点的某个集合．

所谓给定一个点的集合 S，是指对于任何一个点 ω，都可以确定它是不是属于 S．如果是，则记为 $\omega \in S$；如果不是，

则记为 $\omega \in S$. 按照这种定义，单个点也是一个点集. 习惯上还约定不包含任何点的集合也是一个点集，称为空集，记为$\varnothing$.

今后，我们把**事件**定义为样本点的某个集合，称某事件发生当且仅当它所包含的某一个样本点出现.

因此，虽然试验的全部可能结果在试验前就很明确，但是只有到了试验之后，才能确定某一特定的事件是否发生.

我们把样本空间 Ω 也作为一个事件，因为在每次试验中必然出现 Ω 中的某个样本点，也即 Ω 必然发生，所以常称 Ω 为**必然事件**. 类似地，我们把空集$\varnothing$也作为一个事件，它在每次试验中都不会发生，称为**不可能事件**.

必然事件 Ω 在试验中必然发生. 相反地，不可能事件在任何试验中不可能发生，必然事件与不可能事件可以说不是随机事件，但为了今后研究的方便，我们还是把必然事件与不可能事件作为随机事件的两个极端情形来统一处理.

三、事件的运算

在一个样本空间中显然可以定义不止一个事件. 概率论的重要研究课题之一是希望从简单事件的概率推算出复杂的事件的概率. 在实际生活中，往往要求我们同时考察几个在同样条件下的事件以及它们之间的联系. 详细地分析事件之间的关系，不仅帮助我们更深刻地认识事件的本质，而且可以大大简化一些复杂事件的概率计算.

下面就讨论事件间的关系及事件的运算，先讨论两个事件 A 与 B 之间的关系.

若 A 中的每一个样本点都包含在 B 中，则记为 $A \subset B$ 或 $B \supset A$，并称 **A 是 B 的特款**，亦称**事件 B 包含了事件 A**，这时事件 A 发生必然导致事件 B 发生. 例如若以 A 记“来到呼叫不超过 5 个”，以 B 记“来到呼叫不超过 6 个”，则 $A \subset B$. 显然对任何事件 A，必有 $\Omega \supset A \supset \varnothing$.

如果 $A \supset B$ 与 $B \supset A$ 同时成立，则称 **A 与 B 等价**或称 **A 等于 B**，记为 $A=B$，等价的两个事件同时发生，因此可看作是一样的.

对于事件 A，由所有不包含在 A 中的样本点所组成的事件称为 **A 的逆事件**，或称 **A 的对立事件**，记为 $\overline{A}$，$\overline{A}$ 表示 A 不发生. 例如若以 A 表示“来到呼叫不超过 5 个”，则 $\overline{A}$ 表示“来到呼叫超过 5 个”. 显然，若 $\overline{A}$ 是 A 的对立事件，则 A 也是 $\overline{A}$ 的对立事件，即 $\overline{\overline{A}}=A$. 必然事件与不可能事件互为对立事件.

其次，对于事件 A 及事件 B，定义两个新事件：

用 $A \cap B$ 或 AB 表示所有同时属于 A 及 B 的样本点的集合，称它为 **A 与 B 的交**，事件 AB 表示事件 A 与事件 B 同时发生.

用 $A \cup B$ 表示至少属于 A 或 B 中的一个的所有样本点的集合，称它为 **A 与 B 的并**，事件 $A \cup B$ 表示事件 A 或事件 B 或它们二者发生，也即表示事件 A 与事件 B 至少发生一个.

若 $AB=\varnothing$，则表示 A 与 B 不可能同时发生，称 **A 与 B 互不相容**. 样本点是互不相容的.

以后，对于互不相容事件 A 与 B，我们称它们的并为**和**，并记作 $A+B$.

用 $A-B$ 表示包含在 A 中而不包含在 B 中的样本点全体，称之为 **A 与 B 的差**，事件 $A-B$ 表示事件 A 发生而事件 B 不发生，显然 $A-B=A\overline{B}$.

在进行事件的运算时，关于它们的顺序作如下约定：先进行逆的运算，再进行交的运算，最后才进行并或差的运算.

用上面的记号可以把对立事件之间的关系表述如下：$A \cup \overline{A}=\Omega$，$A \cap \overline{A}=\varnothing$，这也可以作为对立事件的定义. 显然 $\overline{A}=\Omega-A$.

我们用例 5 来说明这些关系. (*) 中 30 种可能结果就是样本点全体，它们构成必然事件 Ω.

A 记第一次摸得黑球，则它由第 5 行及第 6 行的 10 个样本点构成；这时 $\overline{A}$ 表示第一次摸得白球，它由第一行至第四行的 20 个样本点构成. 显然 A 与 $\overline{A}$ 互不相容，而且 $A+\overline{A}=\Omega$.

B 记第二次摸得黑球，它由下列 10 个样本点构成：

(1, 5), (1, 6), (2, 5), (2, 6), (3, 5)

(3, 6), (4, 5), (4, 6), (5, 6), (6, 5)

事件 $A\cup B$ 表示第一次或第二次中至少有一次摸得黑球，它包含下列 18 个样本点：

(1, 5), (1, 6), (2, 5), (2, 6), (3, 5), (3, 6)

(4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4)

(5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5)

事件 AB 表示两次都摸得黑球，它由下列两个样本点构成：(5, 6), (6, 5)，这是 A 与 B 共同包含的样本点. 因此 $C=AB$.

事件 $A-B$ 表示第一次摸得黑球而第二次摸出白球，包含了 (5, 1), (5, 2), (5, 3), (5, 4), (6, 1), (6, 2), (6, 3), (6, 4) 等 8 个样本点.

有时用平面上某正方形中的图形来表示事件间的关系或运算较为直观，这种表示法称为文（Venn）图.

事件 Ω, A, $\overline{A}$, $A\cup B$, AB, $A-B$ 在图 2 中分别以阴影表出. 不难理解，$B\supset A$ 相应于 A 的图形完全包含在 B 的图形中；A 和 B 互不相容，则相应于 A 和 B 的图形不相交.

不难把上面定义推广到多个事件的场合.

例如，对于 n 个事件 $A_1, A_2, \cdots, A_n$，用 $A_1\cup A_2\cup\cdots\cup A_n$ 或 $\bigcup\limits_{i=1}^{n}A_i$ 表示 $A_1, A_2, \cdots, A_n$ 中至少发生一个，称为 $A_1, A_2, \cdots, A_n$ 的并，特别当 $A_1, A_2, \cdots, A_n$ 两两互不相容时，并称为和，记作 $A_1+A_2+\cdots+A_n$ 或 $\sum\limits_{i=1}^{n}A_i$. 用 $A_1A_2\cdots A_n$ 或 $\bigcap\limits_{i=1}^{n}A_i$ 表示 $A_1, A_2, \cdots, A_n$ 同时发生等等.

对于可列个事件的场合，我们定义

$$\bigcup_{i=1}^{\infty}A_i=\lim_{n\to\infty}\bigcup_{i=1}^{n}A_i,\quad \bigcap_{i=1}^{\infty}A_i=\lim_{n\to\infty}\bigcap_{i=1}^{n}A_i$$

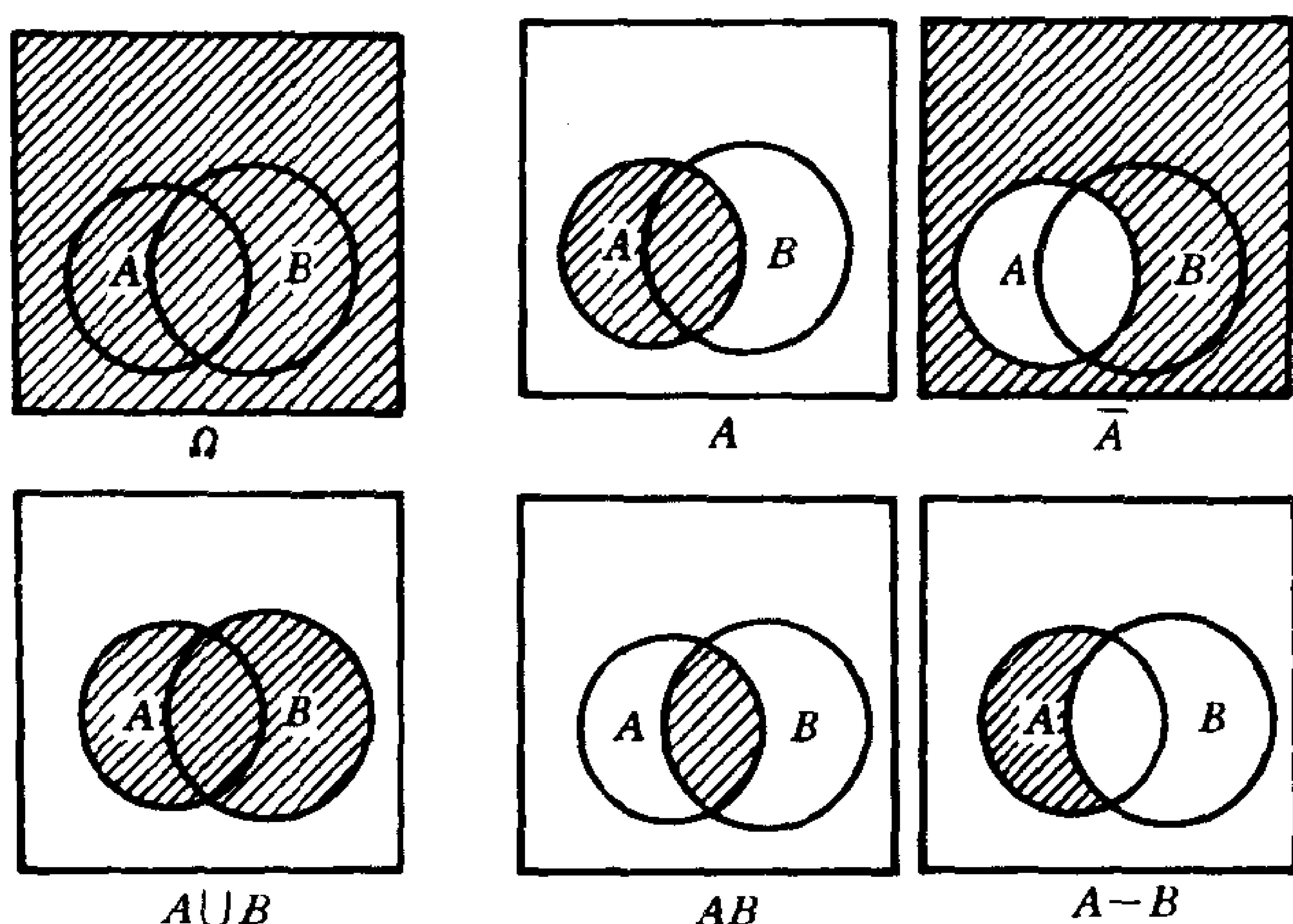


图 2　事件运算

熟悉集合论的读者或许早就发现，事件间的关系及运算与集合论中或布尔(Boole)代数中集合的关系及运算是完全相似的，而且这个相似在建立概率论的严格数学基础时非常重要．不过，我们应该注意另一点，就是要学会用概率论的语言来解释这些关系及运算，并且会用这些运算关系来表示一些事件．

[例 6]　若 A，B，C 是三个事件，则

(1) A 发生而 B 与 C 都不发生可以表示为：$A\overline{B}\overline{C}$ 或 $A-B-C$ 或 $A-(B\cup C)$；

(2) A 与 B 都发生而 C 不发生可以表示为：$AB\overline{C}$ 或 $AB-C$ 或 $AB-ABC$；

(3) 所有这三个事件都发生可以表示为：ABC；

(4) 这三个事件恰好发生一个可以表示为：$A\overline{B}\overline{C}+\overline{A}B\overline{C}+\overline{A}\overline{B}C$；

(5) 这三个事件恰好发生两个可以表示为：$AB\overline{C}+A\overline{B}C+\overline{A}BC$；

(6) 这三个事件中至少发生一个可以表示为：$A\cup B\cup C$ 或

$A\overline{B}\overline{C}+\overline{A}B\overline{C}+\overline{A}\overline{B}C+AB\overline{C}+A\overline{B}C+\overline{A}BC+ABC$；

对于事件的运算成立下列关系式，它们的证明留给读者.

(1) 交换律：$A\cup B=B\cup A$，$AB=BA$；

(2) 结合律：$(A\cup B)\cup C=A\cup(B\cup C)$，$(AB)C=A(BC)$；

(3) 分配律：$(A\cup B)\cap C=AC\cup BC$，

$$(A\cap B)\cup C=(A\cup C)\cap(B\cup C)$$

(4) 德莫根 (De Morgan) 定理：

$$\overline{A_1\cup A_2}=\overline{A}_1\cap\overline{A}_2,\quad \overline{A_1\cap A_2}=\overline{A}_1\cup\overline{A}_2$$

对于 n 个事件，甚至对于可列个事件，德莫根定理也成立.

四、有限样本空间

我们先考虑只有有限个样本点的样本空间，这种样本空间称为**有限样本空间**. 这是最简单的样本空间，研究它有助于深入研究更为复杂的样本空间.

若 Ω 是有限样本空间，其样本点为 E_1，E_2，…，E_n，在这种场合可以把 Ω 的任何子集都当作事件. 在这种样本空间中引进概率，只要对每个样本点 E_i，给定一个数与它对应，此数称为 E_i 的概率，并记之为 $P(E_i)$，它是非负的，而且满足

$$P(E_1)+P(E_2)+\cdots+P(E_n)=1$$

这样，我们对样本点定义了概率，用它来度量每个样本点出现的可能性的大小. 由此出发，我们不难定义更为一般的事件的概率.

定义 1.2.1 任何事件 A 的概率 $P(A)$ 是 A 中各样本点的概率之和.

按照这个定义，显然有 $P(\Omega)=1$，$0\leqslant P(A)\leqslant 1$.

如在例 5 中，若定义每个样本点出现的概率均为 $\frac{1}{30}$（这相当于假定各个球外形完全一样，并且摸球是随机的，各个球被摸到的机会均等），则得 $P(A)=\frac{10}{30}$，$P(B)=\frac{10}{30}$，$P(C)=\frac{2}{30}$，$P(A\cup B)=\frac{18}{30}$，$P(A-B)=\frac{8}{30}$等等.

我们将在下一节研究一种特殊的有限样本空间.

把上面做法推广到有可列个样本点的样本空间是不难的，这种空间称为**离散样本空间**. 但是当把上面做法推广到不可列个样本点的场合，则会遇到实质性的困难，对于这种一般场合的讨论，以后将逐渐展开.

§3. 古 典 概 型

一、模型与计算公式

在讨论一般随机现象之前，我们先讨论一类最简单的随机现象. 这种随机现象具有下列两个特征：

(1) 在试验中它的全部可能结果只有有限个，譬如为 n 个，记为 E_1，E_2，…，E_n，而且这些事件是两两互不相容的；

(2) 事件 E_1，E_2，…，E_n 的发生或出现是等可能的，即它们发生的概率都一样.

这类随机现象在概率论发展初期即被注意，许多最初的概率论结果也是对它作出的，一般把这类随机现象的数学模型称为**古典概型**. 古典概型在概率论中占有相当重要的地位，一方面，由于它简单，对它的讨论有助于直观地理解概率论的许多基本概念，因此，我们常从讨论古典概型开始引入新的概念；另一方面，古典概型概率的计算在产品质量抽样检查等实际问题以及理论物理的研究中都有重要应用.

显然，古典概型是有限样本空间的一种特例. 可以选 $\Omega=\{E_1,E_2,\cdots,E_n\}$ 作为样本空间，而且此时应有

$$P(E_1)=P(E_2)=\cdots=P(E_n)=\frac{1}{n}$$

对于任何事件 A，它总可以表示为样本点之和，例如 $A=E_{i_1}+E_{i_2}+\cdots+E_{i_m}$，因此由事件概率的定义

$$P(A)=P(E_{i_1})+P(E_{i_2})+\cdots+P(E_{i_m})$$

$$=\frac{1}{n}+\frac{1}{n}+\cdots+\frac{1}{n}=\frac{m}{n} \tag{1.3.1}$$

所以在古典概型中，事件 A 的概率是一个分数，其分母是样本点的总数 n，而分子是事件 A 中所包含的样本点的个数 m，由于 E_{i_1}，E_{i_2}，…，E_{i_m}的出现必导致 A 的出现，即它们的出现对 A 的出现“有利”，因此习惯上常称 E_{i_1}，E_{i_2}，…，E_{i_m}是 A 的“有利场合”，这样，

$$P(A)=\frac{m}{n}=\frac{A\text{ 的有利场合的数目}}{\text{样本点总数}} \tag{1.3.2}$$

法国数学家拉普拉斯（Laplace）在 1812 年把上式作为概率的一般定义．现在通常称它为概率的古典定义，因为它只适用于古典概型场合．

古典概型有着多方面应用，产品抽样检查就是其中之一．

产品抽样检查的技术，在各个生产部门中被广泛采用．许多大工厂产量很高，每天生产的产品数以万计，对这些产品的质量如果要进行全面的逐件检验通常是不可能的或是不经济的；另外，在有些情况下，产品的检验方法带有破坏性（如电灯泡寿命检验和棉纱强度试验），这样，最适宜的检验方法是采用抽样检查，即从产品中随机地抽出若干件来检验，并根据检验结果来判断整批产品的质量．

关于产品的质量，可以有多种多样的衡量标准，例如可能要考虑产品的某种形状或尺寸，或把产品分成若干等级，我们先考虑最简单的情形，即把产品分成合格品（好品）与次品（废品）两个类型的场合．

假如产品的好坏从外形上看不出来，而且我们又是随机抽样，那么任何一件产品被抽到的可能性都一样，这正是古典概型．

有一个口袋，内装 a 只黑球，b 只白球，它们除颜色不同外，外形完全一样（以后若非特别声明，均作此假定）．这样一来，当我们从袋子中任意摸出一球时，这 $a+b$ 只球中的任意一只被摸到的可能性都一样．

若把黑球作为废品，白球作为好品，则这个摸球模型就可以描述产品抽样. 假如产品分为更多等级，例如一等品，二等品，三等品，等外品等等，则可用装有多种颜色的球的口袋的摸球模型来描述.

这种模型化的方法能使问题更清楚，更容易看出其随机性本质而不致被个别情况下的具体属性所蒙蔽. 不仅如此，这种抽象化的模型带有普遍性，它还可以描述许多别的具体问题，从而有着多方面应用. 例如种水稻地块的调查，某种疾病的抽查等都能用这个模型.

事实上，古典概型的大部分问题都能形象化地用摸球模型来描述. 以后我们经常研究摸球模型，意义即在于此.

前节例 5 及其有关概率的计算是古典概型的一个例子，但并不是所有古典概型的事件的概率计算都这么容易. 事实上，古典概型中许多概率的计算相当困难而富有技巧. 计算的要点是给定样本点，并计算它的总数，而后再计算有利场合的数目. 在这些计算中，经常要用到一些排列与组合公式.

二、基本的组合分析公式

1. 全部组合分析公式的推导基于下列两条原理：

乘法原理　若进行 A_1 过程有 n_1 种方法，进行 A_2 过程有 n_2 种方法，则进行 A_1 过程后再接着进行 A_2 过程共有 $n_1 \times n_2$ 种方法（图 3）.

加法原理　若进行 A_1 过程有 n_1 种方法，进行 A_2 过程有 n_2 种方法，假定 A_1 过程与 A_2 过程是并行的，则进行过程 A_1 或过程 A_2 的方法共有 n_1+n_2 种（图 4）.

显然这二条原理可以拓广到多个过程的场合.

2. 排列：

从包含有 n 个元素的总体中取出 r 个来进行排列，这时既要考虑到取出的元素也要顾及其取出顺序.

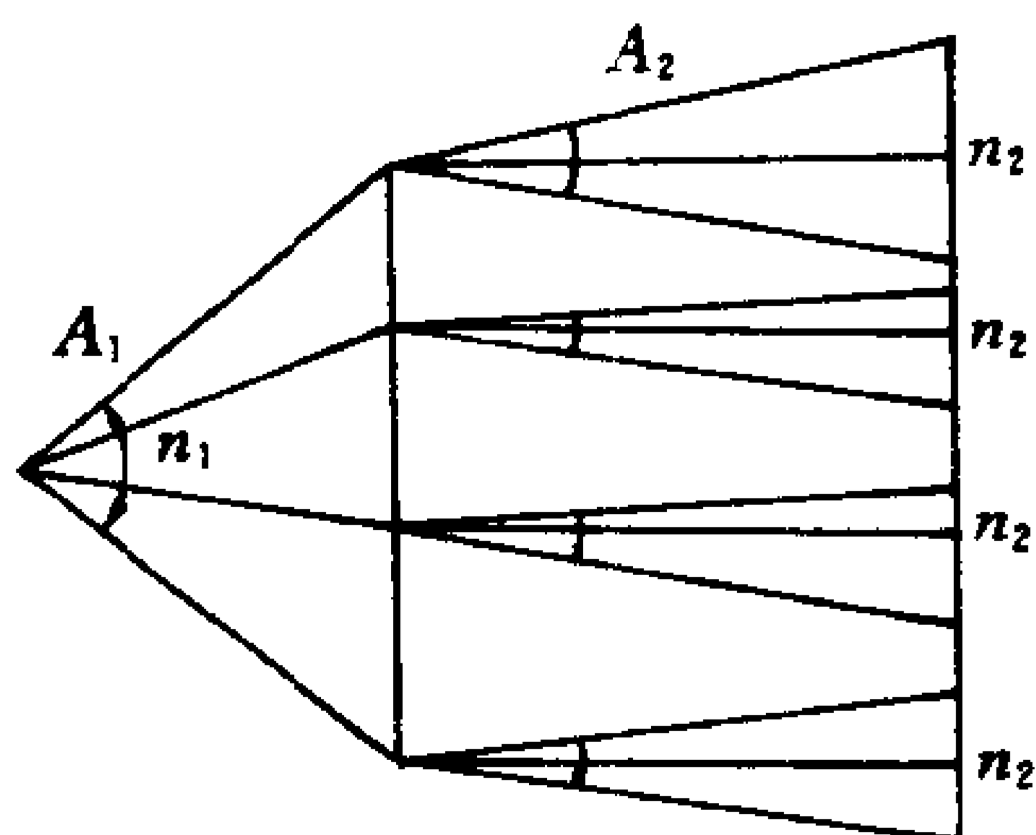


图 3

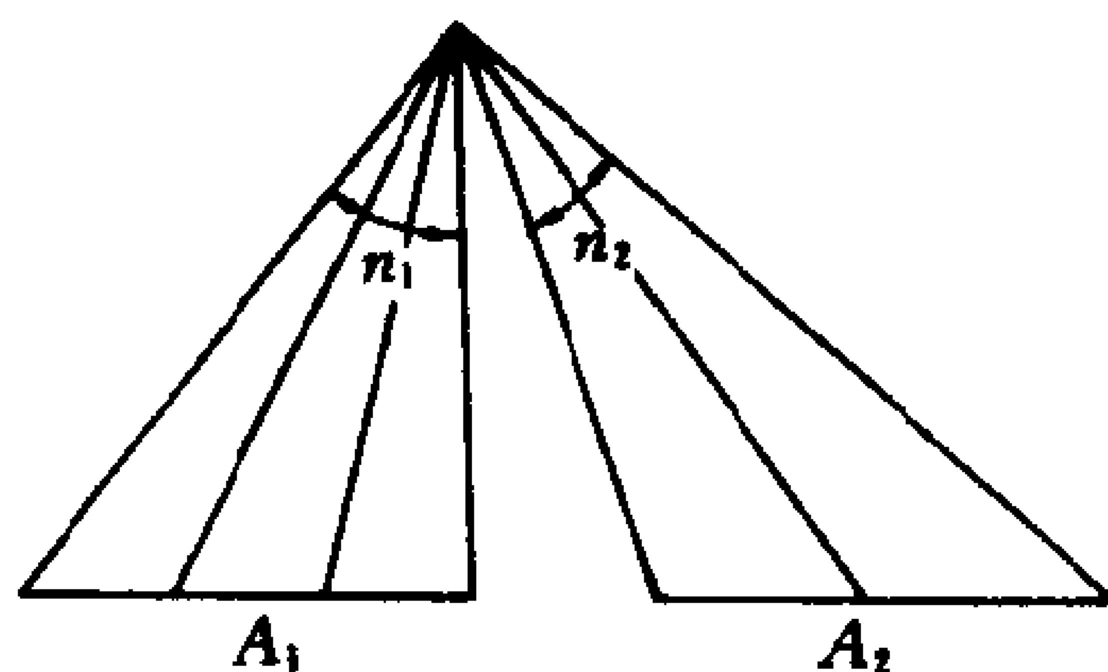


图 4

这种排列可分为两类：第一种是有放回的选取，这时每次选取都是在全体元素中进行，同一元素可被重复选中；另一种是不放回选取，这时一个元素一旦被取出便立刻从总体中除去，因此每个元素至多被选中一次，在后一种情况，必有 $r \leqslant n$.

（1）在有放回选取中，从 n 个元素中取出 r 个元素进行排列，这种排列称为**有重复的排列**，其总数共有 n^r 种.

（2）在不放回选取中，从 n 个元素中取出 r 个元素进行排列，其总数为

$$A_n^r = n(n-1)(n-2)\cdots(n-r+1)$$

这种排列称为**选排列**. 特别当 $r=n$ 时，称为**全排列**.

（3）n 个元素的全排列数为

$$P_n = n(n-1)\cdots 3 \cdot 2 \cdot 1 = n!$$

3. 组合：

（1）从 n 个元素中取出 r 个元素而不考虑其顺序，称为**组合**，其总数为

$$C_n^r = \binom{n}{r} = \frac{A_n^r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}$$

这里 $\binom{n}{r}$ 是二项展开式的系数，$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$.

（2）若 $r_1 + r_2 + \cdots + r_k = n$，把 n 个不同的元素分成 k 个部分，

第一部分 r_1 个，第二部分 r_2 个，……第 k 部分 r_k 个，则不同的分法有

$$\frac{n!}{r_1!\ r_2!\ \cdots r_k!} \tag{1.3.3}$$

种，上式中的数称为**多项系数**，因为它是 $(x_1+x_2+\cdots+x_k)^n$ 展开式中 $x_1^{r_1}x_2^{r_2}\cdots x_k^{r_k}$ 的系数，当 $k=2$ 时，即为组合数.

(3) 若 n 个元素中有 n_1 个带足标"1"，n_2 个带足标"2"，……n_k 个带足标"k"，且 $n_1+n_2+\cdots+n_k=n$，从这 n 个元素中取出 r 个，使得带有足标"i"的元素有 r_i 个 $(1\leqslant i\leqslant k)$，而 $r_1+r_2+\cdots+r_k=r$，这时不同取法的总数为

$$\binom{n_1}{r_1}\binom{n_2}{r_2}\cdots\binom{n_k}{r_k} \tag{1.3.4}$$

这里当然要求 $r_i\leqslant n_i$.

(4) 从 n 个元素中有重复地取 r 个，不计顺序，则不同的取法有

$$\binom{n+r-1}{r}$$

种，这个数称为**有重复组合数**.

4. 一些常用等式：

把排列公式推广到 r 是正整数而 n 是任意实数 x 的场合，有时是需要的，这时记

$$A_x^r=x\ (x-1)\ (x-2)\ \cdots\ (x-r+1)$$

同样定义

$$\binom{x}{r}=\frac{A_x^r}{r!}=\frac{x\ (x-1)\ (x-2)\ \cdots\ (x-r+1)}{r!}$$

及 $0!=1$，$\binom{x}{0}=1$.

对于正整数 n，若 $r>n$，则

$$\binom{n}{r}=0$$

这样一来，二项系数有性质

$$\binom{n}{k}=\binom{n}{n-k}$$

$$\binom{-a}{k}=(-1)^k\binom{a+k-1}{k} \tag{1.3.5}$$

由于

$$(1+x)^n=\sum_{r=0}^{n}\binom{n}{r}x^r$$

故

$$\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{n}=2^n$$

利用幂级数乘法又可以证明

$$\binom{a}{0}\binom{b}{n}+\binom{a}{1}\binom{b}{n-1}+\cdots+\binom{a}{n}\binom{b}{0}=\binom{a+b}{n} \tag{1.3.6}$$

特别地

$$\binom{n}{0}\binom{n}{n}+\binom{n}{1}\binom{n}{n-1}+\cdots+\binom{n}{n}\binom{n}{0}=\binom{2n}{n}$$

即

$$\binom{n}{0}^2+\binom{n}{1}^2+\cdots+\binom{n}{n}^2=\binom{2n}{n} \tag{1.3.7}$$

三、概率直接计算的例子

[例 1] 一部四册的文集按任意次序放到书架上去，问各册自右向左或自左向右恰成 1，2，3，4 的顺序的概率是多少？

[解] 若以 a,b,c,d 分别表示自左向右排列的书的册号，则上述文集放置的方式可与向量 (a, b, c, d) 建立一一对应，因为 a，b，c，d 取值于 1，2，3，4，因此这种向量的总数相当于 4 个元素的全排列数 $4!=24$，由于文集按“任意的”次序放到书架上去，因此这 24 种排列中出现任意一种的可能性都相同，这是古典

概型概率，其有利场合有 2 种，即自左向右或自右向左成 1，2，3，4 顺序，因此所求概率为$\frac{2}{24}=\frac{1}{12}$.

［例 2］ 有10 个电阻，其电阻值分别为 1Ω，2Ω，…，10Ω，从中取出三个，要求取出的三个电阻，一个小于 5Ω，一个等于 5Ω，另一个大于 5Ω，问取一次就能达到要求的概率.

［解］ 把从 10 个电阻中取出 3 个的各种可能取法作为样本点全体，这是古典概型，其总数为$\binom{10}{3}$，有利场合为$\binom{4}{1}\binom{1}{1}\binom{5}{1}$，故所求概率为

$$P=\frac{\binom{4}{1}\binom{1}{1}\binom{5}{1}}{\binom{10}{3}}=\frac{1}{6}$$

［例 3］ 甲有 $n+1$ 个硬币，乙有 n 个硬币，双方投掷之后进行比较，求甲掷出的正面数比乙掷出的正面数多的概率.

［解］ 这个问题初看非经繁复计算难求答案，但是若充分利用它特有的对称性并选择适当的样本空间，则能迅速求解.

若以 A 记"甲的正面数＞乙的正面数"，则 $\overline{A}$ 表示"甲的正面数≤乙的正面数"，但是考虑到甲比乙多一个硬币这一特殊情况，则 $\overline{A}$ 又表示"甲的反面数＞乙的反面数". 再由硬币的对称性，显然 $P(A)=P(\overline{A})$. 因此我们构造只有两个样本点的样本空间 $\Omega=\{A, \overline{A}\}$，便可轻而易举地写出答案 $P(A)=\frac{1}{2}$.

［例 4］ 设有 n 个球，每个球都能以同样的概率$\frac{1}{N}$落到 N 个格子（$N\geqslant n$）的每一个格子中，试求：

（1）某指定的 n 个格子中各有一个球的概率；

（2）任何 n 个格子中各有一个球的概率.

［解］ 这是一个古典概型问题，由于每个球可落入 N 个格子中的任一个，所以 n 个球在 N 个格子中的分布相当于从 N 个元素中选取 n 个进行有重复的排列，故共有 N^n 种可能分布.

在第一个问题中，有利场合相当于 n 个球在那指定的 n 个格子中全排列，总数为 $n!$，因而所求概率为

$$P_1=\frac{n!}{N^n}$$

在第二个问题中，n 个格子可以任意，即可以从 N 个格子中任意选出 n 个来，这种选法共有 $\binom{N}{n}$ 种，对于每种选定的 n 个格子，有利场合正如第一个问题一样为 $n!$，故所求概率为

$$P_2=\frac{\binom{N}{n}n!}{N^n}=\frac{N!}{N^n\ (N-n)!}$$

这个例子是古典概型中一个很典型的问题，不少实际问题可以归结为它.

例如，若把球解释为粒子，把格子解释为相空间中的小区域，则这个问题便相应于统计物理学中的 Maxwell-Boltzmann 统计.

概率论历史上有一个颇为有名的问题：要求参加某次集会的 n 个人中没有两个人生日相同的概率. 若把 n 个人看作上面问题中的 n 个球，而把一年的 365 天作为格子，则 $N=365$，这时 P_2 就给出所求的概率. 例如当 $n=40$ 时，$P_2=0.109$，这个概率已经相当小；而当 $n=50$ 时，$P_2=0.03$；进一步当 $n=55$ 时，P_2 之值只有 0.01，这实在是出乎意料地小. 总之，投球问题中球相遇的概率比预料的大得多，这种意外在研究随机现象中时常遇见，也算是随机现象的特性之一吧！

四、抽签与顺序无关

抽签是人为地引进随机性的一个最简单的例子，不单在体育比赛中广泛采用，而且在日常生活中也时常可见. 关于抽签所体现的公平性，即结果虽然不同但机会却是均等的这种概念，只有通过概率论才能清楚阐明.

进一步可以说，抽签是随机化方法的一种特例，随机化方法

在抽样调查和试验设计中都有广泛应用，是人类可以主动利用随机性的一个例证.

下面的摸球问题值得读者仔细推敲.

[例 5] 口袋中有 a 只黑球，b 只白球，它们除颜色不同外，其他方面没有差别，现在把球随机地一只只摸出来，求第 k 次摸出的一只球是黑球的概率（$1\leqslant k\leqslant a+b$）.

[第一种解法] 把 a 只黑球及 b 只白球都看作是不同的（例如设想把它们进行编号），若把摸出的球依次放在排列成一直线的 $a+b$ 个位置上，则可能的排列法相当于把 $a+b$ 个元素进行全排列，总数为 $(a+b)!$，把它们作为样本点全体. 有利场合数为 $a\times(a+b-1)!$，这是因为第 k 次摸得黑球有 a 种取法，而另外 $(a+b-1)$ 次摸球相当于 $a+b-1$ 只球进行全排列，有 $(a+b-1)!$ 种构成法，故所求概率为

$$P_k=\frac{a\times(a+b-1)!}{(a+b)!}=\frac{a}{a+b}$$

这个结果与 k 无关!

细想一下，就会发觉这个结果与我们平常的生活经验是一致的. 例如在体育比赛中进行抽签，对各队机会均等，与抽签的先后次序无关.

[第二种解法] 把 a 只黑球看作是没有区别的，把 b 只白球也看作是没有区别的. 仍把摸出的球依次放在排列成一直线的 $a+b$ 个位置上，因若把 a 只黑球的位置固定下来则其他位置必然是放白球，而黑球的位置可以有 $\binom{a+b}{a}$ 种放法，以这种放法作为样本点. 这时有利场合数为 $\binom{a+b-1}{a-1}$，这是由于第 k 次摸得黑球，这个位置必须放黑球，剩下的黑球可以在 $a+b-1$ 个位置上任取 $a-1$ 个位置，因此共有 $\binom{a+b-1}{a-1}$ 种放法. 所以所求概率为

$$P_k=\frac{\binom{a+b-1}{a-1}}{\binom{a+b}{a}}=\frac{a}{a+b}$$

两种不同的解法答案相同！

注意考察一下两种解法的不同，就会发现主要在于选取的样本空间不同. 在前一种解法中把球看作是“有个性的”，而在后一种解法中则对同色球不加区别，因此在第一种解法中要顾及各黑球间及各白球间的顺序而用排列，第二种解法则不注意顺序而用组合，但最后还是得出了相同的答案.

这种情况的产生并不奇怪，这说明对于同一随机现象，可以用不同的模型来描述，只要方法正确，结论总是一致的. 在这个例子中，第二种解法中的每一个样本点是由第一种解法中的 $a!\cdot b!$ 个样本点合并而成的.

这个例子还告诉我们，在计算样本点总数及有利场合数时，必须对同一个确定的样本空间考虑，因此其中一个考虑顺序，另一个也必须考虑顺序，否则结果一定不正确.

既然同一个随机现象可用不同的样本空间来描述，因此对同一个概率也常常有多种不同的求法，我们应逐步训练自己能采用最简便的方法解题，为此熟悉同一问题的多种不同解法是重要的.

例如，对例 5 就存在着多种不同的解法，上面提供的只是比较自然的两种. 注意到在这两种解法中，我们对不同的 k 用的是同一个样本空间，也就是说，我们构造了一个可以描述 $a+b$ 次摸球的样本空间，并利用它一举解决了“第 k（$1\leqslant k\leqslant a+b$）次摸得黑球”这一概率的计算. 假如允许对不同的 k 用不同的样本空间，则我们完全可以构造一个只包含前 k 次试验，甚至只包含第 k 次试验的样本空间，这时也能求得有关概率. 特别是选用最后一种样本空间简直马上可以看出正确答案，不过这种做法对初学者或许不那么容易理解.

五、二项分布与超几何分布

产品抽样检验有两类，即有放回抽样与不放回抽样. 在有放回抽样中，被抽出的产品检验后仍放回产品堆中，再抽第二次，因此这件产品以后仍然可能再次被抽到. 更常用的是第二种方法，即不放回的抽样方法，这时被抽到的产品不再放回，因而以后不会再被抽到. 与此相应的，我们的摸球模型也假定为有放回与不放回摸球两种，这两个情形得到的结果是不同的.

下面是古典概型概率计算中一个典型问题，它有着多方面应用，特别在产品检验方面起很大作用.

［例 6］ 如果某批产品中有 a 件次品 b 件好品，我们采用有放回及不放回抽样方式从中抽 n 件产品，问正好有 k 件是次品的概率各是多少？

所求的概率显然与抽样方式有关，下面我们分别来讨论.

［有放回抽样场合］ 把 $a+b$ 件产品进行编号，有放回抽 n 次，把可能的重复排列全体作为样本点，总数为 $(a+b)^n$，其中有利场合（即次品正好出现 k 次）的数目是 $\binom{n}{k}a^kb^{n-k}$，故所求概率为

$$b_k=\frac{\binom{n}{k}a^kb^{n-k}}{(a+b)^n}=\binom{n}{k}\left(\frac{a}{a+b}\right)^k\left(\frac{b}{a+b}\right)^{n-k} \qquad (1.3.8)$$

b_k 是二项式 $\left(\frac{a}{a+b}+\frac{b}{a+b}\right)^n$ 展开式的一般项，上述概率称为**二项分布**（分布一词的意义将在第三章阐明）. 关于二项分布更一般的讨论在以后各章陆续进行.

［不放回抽样场合］ 从 $a+b$ 件产品中取出 n 件产品的可能组合全体作为样本点，总数为 $\binom{a+b}{n}$，有利场合数为 $\binom{a}{k}\binom{b}{n-k}$，故所求概率为

$$h_k=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}} \tag{1.3.9}$$

这个概率称为**超几何分布**.

从直观上看，当产品总数很大而抽样数不大时，采用有放回抽样与采用不放回抽样，差别应该不大.

事实上，因为

$$h_k=\frac{\binom{a}{k}\binom{b}{n-k}}{\binom{a+b}{n}}=\frac{\frac{A_a^k}{k!}\cdot\frac{A_b^{n-k}}{(n-k)!}}{\frac{A_{a+b}^n}{n!}}=\binom{n}{k}\frac{A_a^kA_b^{n-k}}{A_{a+b}^n}$$

$$=\binom{n}{k}\frac{a^kb^{n-k}}{(a+b)^n}\cdot\frac{\frac{A_a^k}{a^k}\cdot\frac{A_b^{n-k}}{b^{n-k}}}{\frac{A_{a+b}^n}{(a+b)^n}}$$

而当 k 比 a 小得多，$n-k$ 比 b 小得多时：

$$\frac{\frac{A_a^k}{a^k}\cdot\frac{A_b^{n-k}}{b^{n-k}}}{\frac{A_{a+b}^n}{(a+b)^n}}\approx 1$$

因此

$$h_k\approx b_k$$

在实际工作中，抽样一般都采用不放回方式，因此计算次品数为 k 的概率时应该用超几何分布，但（1.3.9）的数值计算较繁复. 若产品数甚大而抽样数不太大，则可利用上述性质，计算二项分布作为近似值，这时有许多专门表格可查，这样可以大大节省计算工作量.

利用上例结果，马上可以计算下列概率：若一批产品共有 N 件，其中有次品 M（$M<N$）件，今抽取 n 件，则其中恰有 m 件



次品的概率是

$$P_m=\frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}},\quad \begin{array}{l}0\leqslant n\leqslant N,\ 0\leqslant m\leqslant M\\ 0\leqslant n-m\leqslant N-M\end{array} \tag{1.3.10}$$

这是超几何分布的另一种常见形式.

学到这里，细心的读者可能会发觉这样一个矛盾，在我们前面讨论中都假定产品中的次品数已知，然后根据它来计算种种概率，而在实际问题中，情况恰恰相反，次品数是未知的，并且正是我们希望通过抽样检验来确定的.

这个矛盾可通过下面办法来解决.

不难理解，抽出来的样本的质量情况在某种程度上反映了整批产品的质量情况，例如，如果整批产品中次品很多，则抽查的样本中含有次品的可能性就相当大；反之，若产品中极少次品，则从中抽查一、两件产品而得到次品的可能性就很小，因而样本中所含次品数的多少就为我们估计整批产品中的次品数提供了某种根据. 例如为了确定某批产品的次品率，通常采用的方法是从这批产品中抽若干件产品作为样本来检验，并用样本的次品率来估计整批产品的次品率. 关于这个课题的研究，构成了数理统计的重要内容.

由于抽样带有随机性，因而不同的抽样可能得到不同的结果，所以我们有必要对各种结果出现的可能性大小进行讨论，这为我们根据样本情况推断整批产品情况提供了理论依据，这种研究是概率论的任务. 从这里也看出，概率论与数理统计有着很密切的联系.

[例 7]　从某鱼池中捕得 1200 条鱼，做了红色的记号之后再放回池中，经过适当的时间后，再从池中捕 1000 条鱼，数出其中有红色记号的鱼的数目，共有 100 条，试估计鱼池中共有多少条鱼？

[解]　设池中共有 n 条鱼，n 未知，是我们要估计的. 更一般

地，设第一次捕得的鱼有 n_1 条（针对本例，$n_1=1200$），第二次捕得 r 条（针对本例，$r=1000$），而其中有记号的有 k 条（针对本例，$k=100$）.

现在，在第二次捕鱼中有 k 条有记号鱼的概率按超几何分布（1.3.10）给出：

$$p_k(n)=\binom{n_1}{k}\binom{n-n_1}{r-k}\bigg/\binom{n}{r}$$

因为实际上在 r 条鱼中有 k 条有记号，因此我们求 n 使得上式概率达到最大，并把这个数值作为池中鱼数的估计.

由于

$$\frac{p_k(n)}{p_k(n-1)}=\frac{(n-n_1)\ (n-r)}{(n-n_1-r+k)\ n}=\frac{n^2-nn_1-nr+n_1r}{n^2-nn_1-nr+nk}=\rho$$

因此当 $nk<n_1r$ 时，$\rho>1$，而当 $nk>n_1r$ 时，$\rho<1$，即 $p_k(n)$ 当 $n<\frac{n_1r}{k}$ 时是 n 的增函数，而当 $n>\frac{n_1r}{k}$ 时是 n 的减函数，所以当 n 等于

$$\hat{n}=\left[\frac{n_1r}{k}\right]$$

时，$p_k(n)$ 达到最大值. 这样我们把 $\left[\frac{n_1r}{k}\right]$ 作为鱼池中鱼总数 n 的估计量.

在这个例子中：

$$\hat{n}=\frac{1200\times1000}{100}=12000$$

这里解题的基本思路是把概率 $p_k(n)$ 看作未知参数 n 的函数，再通过求其最大值而得到 n 的估计，这是数理统计中著名的**最大似然估计法**.

六、概率的基本性质

古典概型中，若样本空间为 $\Omega=\{E_1, E_2, \cdots, E_n\}$，分别以

$n(A)$, $n(B)$, $n(A\cup B)$ 及 $n(AB)$ 记事件 A，B，$A\cup B$ 及 AB 所包含的样本点数，则显然成立公式

$$n(A\cup B)=n(A)+n(B)-n(AB)$$

两边除以 n，即得概率的等式

$$P(A\cup B)=P(A)+P(B)-P(AB)$$

以后将会看到，这是普遍成立的一个等式. 如果这里的 A 与 B 不相容，即 $AB=\varnothing$，立刻得到

$$P(A+B)=P(A)+P(B) \qquad (1.3.11)$$

不难把这个结果推广到有限个事件的场合，即若 A_1，A_2，…，A_m 是 m 个两两互不相容的事件，则

$$P(A_1+A_2+\cdots+A_m)=P(A_1)+P(A_2)+\cdots+P(A_m) \qquad (1.3.12)$$

这个结论可借助（1.3.11）用归纳法证明.

系 $P(\overline{A})=1-P(A)$. (1.3.13)

［证］ 由于 $A+\overline{A}=\Omega$，$A\overline{A}=\varnothing$ 所以

$$P(\Omega)=P(A+\overline{A})=P(A)+P(\overline{A})=1$$

因此

$$P(\overline{A})=1-P(A)$$

关系式（1.3.13）很简单，意义也很明确，但相当有用，它为计算某些事件的概率提供了很大方便.

［例 8］(德·梅尔问题） 一颗骰子投 4 次至少得到一个六点与两颗骰子投 24 次至少得到一个双六，这两件事中哪一件有更多的机会遇到?

［解］ 以 A 表示一颗骰子投 4 次至少得到一个六点这一事件，为求 $P(A)$，在这种场合最方便的方法是利用(1.3.13)式先求 $P(\overline{A})$，这时 $\overline{A}$ 表示投一颗骰子 4 次都没有出现六点，因此不难得出

$$P(\overline{A})=\frac{5^4}{6^4}$$

从而得到

$$p_1=P(A)=1-\left(\frac{5}{6}\right)^4=0.5177$$

若以 B 表示两颗骰子投 24 次至少得到一个双六这一事件，则用同样的方法可以求得

$$p_2=P(B)=1-\left(\frac{35}{36}\right)^{24}=0.4914$$

这样一来，我们知道前者的机会大于$\frac{1}{2}$，而后者的机会小于$\frac{1}{2}$.

这个问题在概率论发展史上颇有名气，因为它是德·梅尔向巴斯卡（Pascal）提出的问题之一．正是这些问题导致了巴斯卡的研究和他与费马（Fermat）的著名通信．前已指出，他们的研究标志着概率论的诞生．

［例 9］ 一口袋中装有 $N-1$ 只黑球及 1 只白球，每次从袋中随机地摸出一球，并换入一只黑球，这样继续下去，问第 k 次摸球时摸到黑球的概率是多少？

［解］ 若以 A 表示第 k 次摸到黑球这一事件，则 $\overline{A}$ 表示第 k 次摸到白球．现在计算 $P(\overline{A})$.

因为袋中只有一只白球，而每次摸出球总是换入黑球，故为了在第 k 次摸到白球，则前面的 $k-1$ 次一定不能摸到白球．因此 $\overline{A}$ 等价于下列事件：在前 $k-1$ 次摸球时都摸出黑球而第 k 次摸出白球，这一事件的概率为

$$\frac{(N-1)^{k-1}\cdot 1}{N^k}=\left(1-\frac{1}{N}\right)^{k-1}\cdot\frac{1}{N}$$

这样

$$P(A)=1-P(\overline{A})=1-\left(1-\frac{1}{N}\right)^{k-1}\cdot\frac{1}{N}$$

公式（1.3.13）的应用使这个题目很快获得解决，假如直接计算 $P(A)$则困难得多．从公式(1.3.13)中容易看出，只要 $P(A)$或 $P(\overline{A})$中的任何一个知道了，则可以求得另一个．在不同问题中，有的求 $P(A)$容易，求 $P(\overline{A})$ 困难；有的正好相反．利用

(1.3.13)，我们就可以先求容易的一个，再去求另一个.

从古典概型的概率研究中，我们发现概率有下面三个基本性质：

(i) 对于任何事件 A，$P(A) \geqslant 0$；

(ii) $P(\Omega) = 1$；

(iii) 若 A_1，A_2，…，A_m 两两互不相容，则

$$P(A_1+A_2+\cdots+A_m)=P(A_1)+P(A_2)+\cdots+P(A_m)$$

第一个性质称为概率的**非负性**，第二个性质称为概率的**规范性**，第三个性质称为概率的（**有限**）**可加性**.

§4. 几 何 概 率

一、例子与计算公式

在古典概型中利用等可能性的概念，成功地计算了某一类问题的概率；不过，古典概型要求可能场合的总数必须有限，因此历史上有不少人企图把这种做法推广到有无限多结果而又有某种等可能性的场合. 这类问题一般可以通过几何方法来求解.

先从几个简单的例子开始.

[例 1] 某人午觉醒来，发觉表停了，他打开收音机，想听电台报时，求他等待的时间短于 10 分钟的概率.

[例 2] 如果在一个 5 万平方公里的海域里有表面积达 40 平方公里的大陆架贮藏着石油，假如在这海域里随意选定一点钻探，问钻到石油的概率是多少？

[例 3] 在400 毫升自来水中有一个大肠杆菌，今从中随机取出 2 毫升水样放到显微镜下观察，求发现大肠杆菌的概率.

一种相当自然的答案是认为例 1 所求的概率等于$\frac{1}{6}$，例 2 中钻到石油的概率等于$\frac{8}{10000}$，而例 3 所求的概率等于$\frac{1}{200}$. 在求这些概率时，我们事实上是利用了几何的方法，并假定了某种等可能性.

在例 1 中，因为电台每小时报时一次，我们自然认为这个人打开收音机时处于两次报时之间，例如（13：00，14：00），而且取各点的可能性一样，要遇到等待时间短于 10 分钟，只有当他打开收音机的时间正好处于 13：50 至 14：00 之间才有可能，相应的概率是$\frac{10}{60}=\frac{1}{6}$.

在例 2 中，由于选点的随机性，可以认为该海域中各点被选中的可能性是一样的，因而所求概率自然认为等于贮油海域的面积与整个海域面积之比，即等于$\frac{40}{50000}$.

同样地，例 3 中由于取水样的随机性，所求概率等于水样的体积与总体积之比$\frac{2}{400}$.

总之，在这类问题中，试验的可能结果是某区域 Ω 中的一个点. 这个区域可以是一维的，也可以是二维的，还可以是三维的，甚至可以是 n 维的，这时不管是可能结果全体或是我们所感兴趣的结果都是无限的. 因而等可能性是通过下列方式来赋予意义的：落在某区域 g 的概率与区域 g 的测度（长度、面积、体积等等）成正比并且与其位置及形状无关.

因此，若以 A_g 记“在区域 Ω 中随机地取一点，而该点落在区域 g 中”这一事件，则其概率定义为

$$P(A_g)=\frac{g\text{ 的测度}}{\Omega\text{ 的测度}} \tag{1.4.1}$$

据此定义，则上述诸例之解是明显的. 下面再举一个例子.

［例 4］（会面问题） 两人相约 7 点到 8 点在某地会面，先到者等候另一人 20 分钟，过时就可离去，试求这两人能会面的概率.

［解］ 以 x，y 分别表示两人到达时刻，则会面的充要条件为

$$|x-y|\leqslant 20$$

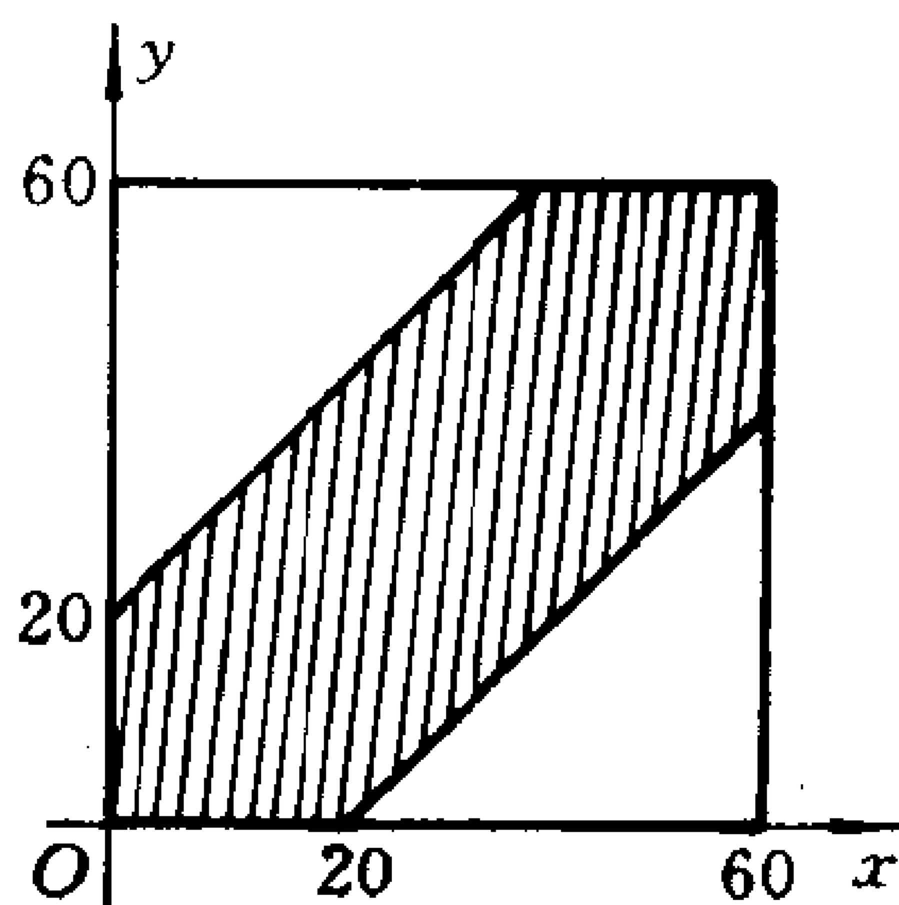


图 5

这是一个几何概率问题，可能的结果全体是边长为 60 的正方形里的点，能会面的点的区域用阴影标出（图 5）. 所求概率为

$$p=\frac{60^2-40^2}{60^2}=\frac{5}{9}$$

二、蒲丰问题

1777 年法国科学家蒲丰提出了下列著名问题，这是几何概率的一个早期例子.

[投针问题] 平面上画着一些平行线，它们之间的距离都等于 a，向此平面任投一长度为 l $(l<a)$ 的针，试求此针与任一平行线相交的概率.

[解] 以 x 表示针的中点到最近的一条平行线的距离，φ 表示针与平行线的交角. 针与平行线的位置关系见图 6.

显然有 $0\leqslant x\leqslant\frac{a}{2}$，$0\leqslant\varphi\leqslant\pi$，以 G 表示边长为 $\frac{a}{2}$ 及 π 的长方形. 为使针与平行线相交，必须 $x\leqslant\frac{l}{2}\sin\varphi$，满足这个关系式的区域记为 g，在图 6 中用阴影表出，所求的概率为

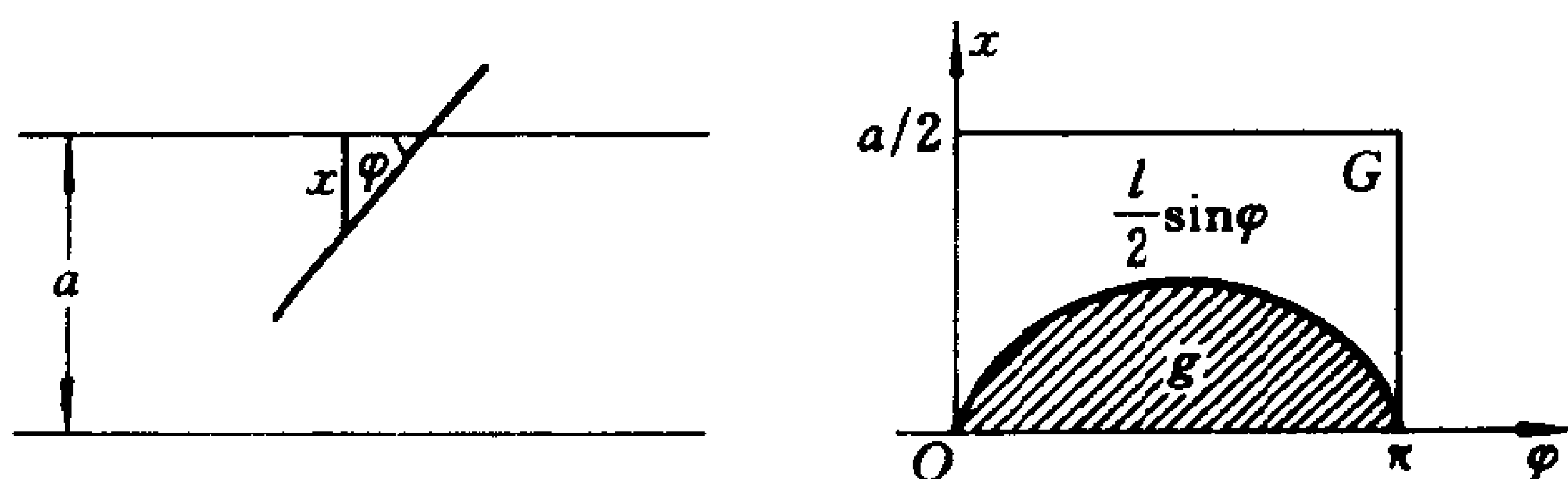


图 6

$$p=\frac{g\text{ 的面积}}{G\text{ 的面积}}=\frac{\frac{1}{2}\int_0^{\pi}l\sin\varphi\,\mathrm{d}\varphi}{\frac{1}{2}a\pi}=\frac{2l}{\pi a}\qquad(1.4.2)$$

由于最后的答案与 π 有关，因此不少人想利用它来计算 π 的数值，其方法是投针 N 次，计算针与线相交的次数 n，再以频率

值$\frac{n}{N}$作为概率 p 之值代入 (1.4.2)，求得

$$\pi=\frac{2lN}{an} \tag{1.4.3}$$

下表[①] 给出了这些试验的有关资料（把 a 折算为 1).

实　验　者	年　份	针　长	投掷次数	相交次数	π 的实验值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
De Morgan，C.	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	$0.8\dot{3}$	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

值得注意的是这里采用的方法：建立一个概率模型，它与某些我们感兴趣的量——这里是常数 π——有关，然后设计适当的随机试验，并通过这个试验的结果来确定这些量.

现在，随着计算机的发展，已按照上述思路建立起一类新的方法，称蒙特卡罗（Monte-Carlo）方法，并取得广泛应用.

三、贝特朗（Bertrand）奇论

几何概率在现代概率概念的发展中曾经起过重大作用. 19 世纪时，不少人相信，只要找到适当的等可能性描述，就可以给概率问题以唯一的解答，然而有人却构造出这样的例子，它包含着几种似乎都同样有理但却互相矛盾的答案，下面就是一个著名的例子.

［贝特朗奇论］　在半径为 1 的圆内随机地取一条弦，问其长超过该圆内接等边三角形的边长 $\sqrt{3}$ 的概率等于多少？

---

① 引自 Gridgeman, N. T. Geometric probability and the number π. Scripta Mathematica，25（1960），183—195.

这是一个几何概率问题，但是基于对术语“随机地”的含义的不同解释，这个问题却存在多种不同答案，下面是其中的三种.

[解法一]　任何弦交圆周两点，不失一般性，先固定其中一点于圆周上，以此点为顶点作一等边三角形，显然只有落入此三角形内的弦才满足要求，这种弦的另一端跑过的弧长为整个圆周的$\frac{1}{3}$，故所求概率等于$\frac{1}{3}$（见图 7 (a)).

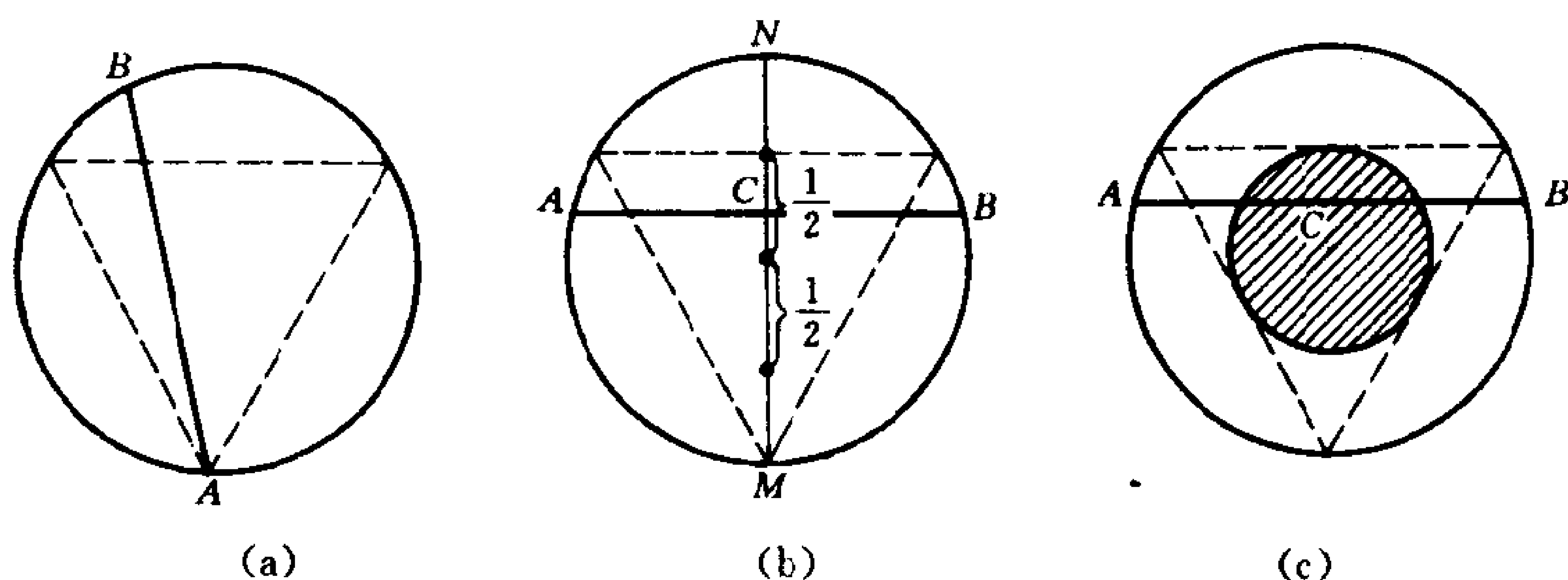


图 7

[解法二]　弦长只跟它与圆心的距离有关，而与方向无关，因此可以假定它垂直于某一直径. 当且仅当它与圆心的距离小于$\frac{1}{2}$时，其长才大于$\sqrt{3}$，因此所求概率为$\frac{1}{2}$（见图 7 (b)).

[解法三]　弦被其中点唯一确定，当且仅当其中点属于半径为$\frac{1}{2}$的同心圆内时，弦长大于$\sqrt{3}$，此小圆面积为大圆面积的$\frac{1}{4}$，故所求概率等于$\frac{1}{4}$（见图 7 (c)).

同一问题有三种不同的答案，细究其原因，发现是在取弦时采用不同的等可能性假定. 在第一种解法中，假定端点在圆周上均匀分布，在第二种解法中则假定弦的中点在直径上均匀分布，而在第三种解法中又假定弦的中点在圆内均匀分布. 这三种答案是针对三种不同的随机试验，对于各自的随机试验而言，它们都是正确的.

因此在使用术语“随机”、“等可能”、“均匀分布”等时，应

明确指明其含义；这又因试验而异.

1899 年贝特朗在巴黎出版《概率论》，书中对几何概率提出了批评，并以生动的实例引起大家的注意. 这种善意的批评，推动了概率论的发展.

由于采用等可能性来定义概率有这种困难，因此后来就选择另外的途径，即在定义概率这一基本概念时只指明概率应具有的基本性质，而把具体概率的给定放在一边. 这样做的好处是能针对不同的随机试验给定适当的概率.

与概率的频率解释及古典概型一样，几何概率的研究对于我们了解应要求概率具有哪些基本性质是很有帮助的.

四、几何概率基本性质

几何概率的定义及计算与几何图形的测度密切相关，因此所考虑的事件应是某种可定义测度的集合. 这类集合的并、交也还应该是事件，甚至对它们的可列次并、交也应有这个要求. 例如考察在 (0，1) 中投一个点的随机试验，若以 A 记该点落入 $\left(0, \frac{1}{2}\right)$ 中这个事件，而以 A_n 记该点落入 $\left[\frac{1}{2^{n+1}}, \frac{1}{2^n}\right)$ 中这一事件，$n=1, 2, \cdots$ 则

$$A = \sum_{n=1}^{\infty} A_n$$

如果假定所投的点落入某区间的概率等于该区间的长度，则 $P(A) = \frac{1}{2}$，而 $P(A_n) = \frac{1}{2^{n+1}}$，这时有

$$P(A) = \sum_{n=1}^{\infty} P(A_n)$$

这里我们遇到了事件及概率的可列运算.

综上所述，几何概率应具有如下性质：

(i) 对任何事件 A，$P(A) \geqslant 0$；

(ii) $P(\Omega) = 1$；

(iii) 若 A_1，A_2，…两两互不相容，则

$$P\left(\sum_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P(A_n) \tag{1.4.4}$$

前两个性质与古典型概率相同，(1.4.4) 则要求对可列个两两互不相容事件成立，这性质称为**可列可加性**.

§5. 概 率 空 间

一、走向概率论公理化结构

到本世纪初，概率论的各个领域已经得到了大量的成果，而且人们对概率论在其他基础学科和工程技术上的应用也出现了越来越大的兴趣. 但是直到那时为止，关于概率论的一些基本概念——例如事件、概率等——却没有明确的定义. 这是一个很大的矛盾，这个矛盾不仅可能导致贝特朗奇论那样的怪现象产生，而且也使许多人对概率的客观含义，甚至概率论结论的可应用性都产生了怀疑. 因此可以说，到那时为止，概率论作为一个数学分支来说，还缺乏严格的理论基础，这就大大妨碍了它的进一步发展.

在概率论发展早期，所研究的随机现象比较简单，大部分可以归入古典概型. 对这种模型，概率的计算可以通过某种等可能性的假设进行，其结果也有相当明确的解释. 这种成功使得人们试图通过给定某种等可能性来定义概率. 于是，由拉普拉斯给出的概率的古典定义在整个19世纪被人们广泛接受. 但是，这种定义的局限性很快也暴露了出来，它既要求试验的可能结果总数有限，又要求某种等可能性，所以它的适用范围有限. 当把这个结果推广到有无限多种可能结果的场合，例如几何概率时，不但适用范围仍然有限，而且还出现了新问题. 总之，对一般的随机现象明确地定义概率及其他基本概念，在那时成了一个突出的问题.

解决这个问题的时机也在不断成熟. 首先是通过对概率论的

两个最基本概念——事件与概率的长期研究，发觉事件的运算与集合的运算完全相似，概率与测度有相同性质，这个事实随着当时在实变函数论中关于勒贝格（Lebesgue）测度和积分的研究以及一般抽象测度和积分理论的发展而日益明确起来.

另外，19 世纪末以来，数学的各个分支广泛流行着一股公理化潮流，这个流派主张把最基本的假定公理化，其他结论则由它们经过演绎导出.

在这种背景下，1933 年，前苏联数学家科尔莫戈罗夫（A. H. Колмогоров）提出了概率论公理化结构，这个结构综合了前人成果，明确定义了基本概念，使概率论成为严谨的数学分支，对近几十年来概率论的迅速发展起了积极作用. 科尔莫戈罗夫的这个理论已被普遍接受，本书的各部分就是在这个结构下展开的.

二、事件域

在公理化结构中，事件不是最基本的概念，它通过更基本的概念——样本点来定义. **样本点**相应于随机试验的结果，我们已在前三节中进行过不少描述. 在古典概型中，它们是可能结果全体，被用来定义一般的事件. 在几何概率中，它们相应于区域 Ω 中的点，也被用来定义一般的事件. 以后，我们把样本点 ω 看作抽象的点，它们的全体构成**样本空间** Ω.

正如 § 2 中所做的那样，我们把事件 A 定义为 Ω 的一个子集，它包含若干样本点，事件 A 发生当且仅当 A 所包含的样本点中有一个发生.

一般并不把 Ω 的一切子集都作为事件，因为这将对给定概率带来困难，譬如在几何概率中，若把不可测集也作为事件，将带来不可克服的困难.

另一方面，又必须把问题中感兴趣的事件都包括进来. 例如若 A 是事件，则应要求 $\overline{A}$ 也是事件；若 A 与 B 是事件，则 $A\cup B$ 及 AB 也应是事件. 当样本空间 Ω 由无限多个点构成时——在几

何概率中就是如此——显然还得考虑可列个事件的并与交. 此外, 把 Ω 及 $\varnothing$ 作为事件有很大方便.

总之, 我们若把事件的全体记为 $\mathscr{F}$, 它是由 Ω 的一些子集构成的集类. 而且为了使讨论便于进行, 还得对 $\mathscr{F}$ 加上某些限制:

(i) $\Omega \in \mathscr{F}$;

(ii) 若 $A \in \mathscr{F}$, 则 $\overline{A} \in \mathscr{F}$;

(iii) 若 $A_n \in \mathscr{F}$, $n=1, 2, \cdots$ 则 $\bigcup_{n=1}^{\infty} A_n \in \mathscr{F}$.

一般地, 称空间 Ω 上满足上述三个要求的集类为 **σ 域**, 亦称 **σ 代数**.

若 $\mathscr{F}$ 为 σ 域, 则由 (i) 及 (ii) 可得

$$\varnothing \in \mathscr{F}$$

此外, 若 $A_n \in \mathscr{F}$, $n=1, 2, \cdots$ 则

$$\bigcap_{n=1}^{\infty} A_n = \overline{\bigcup_{n=1}^{\infty} \overline{A}_n} \in \mathscr{F}$$

$$\bigcup_{n=1}^{k} A_n = A_1 \cup A_2 \cup \cdots \cup A_k \cup \varnothing \cup \varnothing \cdots \in \mathscr{F}$$

$$\bigcap_{n=1}^{k} A_n = A_1 \cap A_2 \cap \cdots \cap A_k \cap \Omega \cap \Omega \cdots \in \mathscr{F}$$

定义 1.5.1 若 $\mathscr{F}$ 是由样本空间 Ω 的一些子集构成的一个 σ 域, 则称它为**事件域** (event field), $\mathscr{F}$ 中的元素称为**事件**, Ω 称为**必然事件**, $\varnothing$ 称为**不可能事件**.

值得指出, 按照这种定义, 样本点并不一定是事件.

下面我们来举一些例子.

[例 1] $\mathscr{F} = \{\varnothing, \Omega\}$, 不难验证 $\mathscr{F}$ 是一个 σ 域, 这时只有必然事件 Ω 与不可能事件 $\varnothing$ 是事件.

[例 2] $\mathscr{F} = \{\varnothing, A, \overline{A}, \Omega\}$. 这时 $\mathscr{F}$ 也是一个 σ 域, Ω, A, $\overline{A}$, $\varnothing$ 是事件.

[例 3] $\Omega = \{\omega_1, \cdots, \omega_n\}$, $\mathscr{F}$ 由 Ω 的一切子集构成, 这时 $\mathscr{F}$ 是一个有限的集合, 共有 2^n 个元素, 不难验证, $\mathscr{F}$ 是一个 σ 域.

［例 4］　对于一般的 Ω，若 $\mathscr{F}$ 由 Ω 的一切子集构成，可以验证 $\mathscr{F}$ 是一个 σ 域.

从上面几个例子中看到，事件域可以很简单，也可以选得十分复杂，这就需要我们根据问题的不同要求来选择适当的事件域.

下面证明：若给定 Ω 的一个非空集类 $\mathscr{G}$，必存在唯一的一个 Ω 中的 σ 域 m（$\mathscr{G}$），具有如下两个性质：(1) 包含 $\mathscr{G}$；(2) 若有其他 σ 域包含 $\mathscr{G}$，则必包含 m（$\mathscr{G}$）. m（$\mathscr{G}$）称为**包含 $\mathscr{G}$ 的最小 σ 域**，亦称**由 $\mathscr{G}$ 产生的 σ 域**.

先证明必存在包含 $\mathscr{G}$ 的 σ 域，显然由 Ω 的一切子集构成的集类包含了 $\mathscr{G}$，由例 4 知此集类是一个 σ 域，因此至少存在 Ω 中的一个 σ 域 m，有 m$\supset\mathscr{G}$.

现在只要取 Ω 中一切包含 $\mathscr{G}$ 的 σ 域之交作为 m（$\mathscr{G}$），则它是具有上述两个性质的 σ 域. 这点留给读者自行验证.

下面介绍两个非常有用的 σ 域.

［一维博雷尔（Borel）点集］　以后我们将以 $\mathbf{R}^1$ 记数直线或实数全体，并称由一切形为 $[a, b)$ 的有界左闭右开区间构成的集类所产生的 σ 域为**一维博雷尔 σ 域**，记之为 $\mathscr{B}_1$，称 $\mathscr{B}_1$ 中的集为**一维博雷尔点集**.

若 x，y 表示任意实数，由于

$$\{x\} = \bigcap_{n=1}^{\infty}\left[x, x+\frac{1}{n}\right)$$

$$(x, y) = [x, y) - \{x\}$$

$$[x, y] = [x, y) + \{y\}$$

$$(x, y] = [x, y) + \{y\} - \{x\}$$

因此 $\mathscr{B}_1$ 中包含一切开区间，闭区间，单个实数，可列个实数，以及由它们经可列次并、交运算而得出的集合. 这是相当大的一个集类，足够把实际问题中感兴趣的点集都包括在内.

［n 维博雷尔点集］　以 $\mathbf{R}^n$ 记 n 维欧几里得（Euclid）空间，可以类似地定义 **n 维博雷尔点集**，它们是由一切 n 维矩形产生的 n

维波雷尔 σ 域. $\mathscr{B}_n$ 中的集合，也可以把 $\mathbf{R}^n$ 中我们感兴趣的点集都包括在内.

三、概率

在公理化结构中，概率是针对事件定义的，即对应于事件域 $\mathscr{F}$ 中的每一个元素 A 有一个实数 $P(A)$ 与之对应，一般把这种从集合到实数的映照称为集合函数. 因此，概率是定义在事件域 $\mathscr{F}$ 上的一个集合函数. 此外，在公理化结构中只规定概率应满足的性质，而不具体给出它的计算公式或计算方法.

概率应有什么性质呢？

因为概率通过频率稳定性与随机试验相联系，因此我们自然想到概率应有与频率类似的性质. 关于频率的性质，我们已在 § 1 中总结为非负性，规范性以及有限可加性.

在古典概型中，概率是通过有利场合数与可能结果总数之比来定义的，它同样具有这三个性质.

在几何概率中，情况也类似，但有一点不同，就是它要求对可列个不相容事件之和有可加性，即可列可加性.

在一般场合，处理可列个事件之和是完全必要的，因此保留这种可列可加性要求看来是合理的.

综上所述，我们如下定义概率.

定义 1.5.2 定义在事件域 $\mathscr{F}$ 上的一个集合函数 P 称为**概率**，如果它满足如下三个条件：

(i) $P(A) \geqslant 0$，对一切 $A \in \mathscr{F}$；

(ii) $P(\Omega) = 1$；

(iii) 若 $A_i \in \mathscr{F}, i = 1,2,\cdots$ 且两两互不相容，则

$$P\Big(\sum_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} P(A_i) \tag{1.5.1}$$

性质 (iii) 称为**可列可加性**或**完全可加性**.

利用概率的基本性质 (i)，(ii)，(iii) 可以推出概率的另外一

些重要性质.

性质 1 不可能事件的概率为 0，即 $P(\varnothing)=0$.

[证明] 因为

$$\Omega=\Omega+\varnothing+\cdots$$

所以

$$P(\Omega)=P(\Omega)+P(\varnothing)+\cdots$$

因此 $P(\varnothing)=0$.

性质 2 概率具有有限可加性. 即若 $A_iA_j=\varnothing\ (i\neq j)$，则

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n) \tag{1.5.2}$$

[证明] 因为

$$A_1+A_2+\cdots+A_n=A_1+A_2+\cdots+A_n+\varnothing+\varnothing+\cdots$$

由可列可加性及性质 1

$$P(A_1+A_2+\cdots+A_n)=P(A_1)+P(A_2)+\cdots+P(A_n)$$

性质 3 对任何事件 A 有

$$P(\overline{A})=1-P(A) \tag{1.5.3}$$

[证明] 因 $A\cup\overline{A}=\Omega$，$A\overline{A}=\varnothing$，故

$$1=P(\Omega)=P(A\cup\overline{A})=P(A)+P(\overline{A})$$

性质 4 如果 $A\supset B$，则

$$P(A-B)=P(A)-P(B) \tag{1.5.4}$$

[证明] 因为当 $A\supset B$ 时，$A=B\cup(A-B)$，又 $B\cap(A-B)=\varnothing$，故 $P(A)=P(B)+P(A-B)$，移项即得结论.

推论 1 如果 $A\supset B$，则 $P(A)\geqslant P(B)$.

由此即知，对任意事件 A，有 $P(A)\leqslant 1$，再注意到概率的非负性，所以成立

$$0\leqslant P(A)\leqslant 1 \tag{1.5.5}$$

性质 5 $P(A\cup B)=P(A)+P(B)-P(AB)$. (1.5.6)

[证明] 因 $A\cup B=A\cup(B-AB)$，而且 $A\cap(B-AB)=\varnothing$，故 $P(A\cup B)=P(A)+P(B-AB)$，又 $AB\subset B$，于是由

性质 4 得到

$$P(A \cup B) = P(A) + P(B) - P(AB)$$

公式（1.5.6）称为**概率的加法公式**，在古典概型场合，我们曾直接加以证明，如今则作为可加性的直接推论，说明在一般场合也成立．由它可得如下推论：

推论 2 （布尔不等式）

$$P(A \cup B) \leqslant P(A) + P(B) \tag{1.5.7}$$

推论 3 （Bonferroni 不等式）

$$P(AB) \geqslant P(A) + P(B) - 1 \tag{1.5.8}$$

利用归纳法不难把这两个不等式推广到 n 个事件的场合．

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant P(A_1) + P(A_2) + \cdots + P(A_n)$$

$$P(A_1A_2\cdots A_n) \geqslant P(A_1) + P(A_2) + \cdots + P(A_n) - (n-1)$$

性质 6 （一般加法公式） 若 $A_1, A_2, \cdots, A_n$ 为 n 个事件，则

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1,\cdots,n} P(A_i) - \sum_{\substack{i<j \\ i,j=1,\cdots,n}} P(A_iA_j) + \sum_{\substack{i<j<k \\ i,j,k=1,\cdots,n}} P(A_iA_jA_k) + \cdots + (-1)^{n-1}P(A_1A_2\cdots A_n) \tag{1.5.9}$$

公式（1.5.9）可以用数学归纳法证明，留给读者作为习题．

当然，概率还有其他性质，不过上述列出的各条是最重要的．它们是以后讨论中经常要用到的基本结果，务必牢记．

值得提醒的是概率的这些重要性质的推导实质上只用到非负性、规范性和有限可加性．

[概率计算的公式法] 利用上面推导的公式来作概率计算，常能使解题思路清晰，计算便捷．

[例 5] 某城有 N 辆卡车，车牌号从 1 到 N，有一个外地人到该城去，把遇到的 n 辆车子的牌号抄下（可能重复抄到某些车牌号），求抄到的最大号码正好为 k 的概率（$1 \leqslant k \leqslant N$）．

［解］　可以看作古典概型问题，即设每辆卡车被遇到的机会相同．若以 A_k 记抄到的最大号码为 k 这一事件，又以 B_k 记抄到的最大号码不超过 k 这一事件，则明显有 $A_k = B_k - B_{k-1}$，而且 $B_k \supset B_{k-1}$，所以由性质 4 知 $P(A_k) = P(B_k) - P(B_{k-1})$，而由直接计算可得 $P(B_k) = \frac{k^n}{N^n}$，因此最后得到

$$P(A_k) = \frac{k^n - (k-1)^n}{N^n}$$

这种方法曾在第二次世界大战中被盟军用来估计敌方的军火生产能力，从被击毁的战车上的出厂号码推测其生产批量，得到相当精确的有用情报．

［例 6］　（匹配问题）　某人写好 n 封信，又写好 n 只信封，然后在黑暗中把每封信放入一只信封中，试求至少有一封信放对的概率．

［解］　若以 A_i 记第 i 封信与信封符合，则所求事件为 $A_1 \cup A_2 \cup \cdots \cup A_n$，所以可以用一般加法公式，不难求得 $P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$，$P(A_iA_j) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}$，$P(A_iA_jA_k) = \frac{1}{n(n-1)(n-2)}$，…，$P(A_1A_2\cdots A_n) = \frac{1}{n!}$．

因此

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \binom{n}{1}\frac{1}{n} - \binom{n}{2}\frac{1}{n(n-1)} + \binom{n}{3}\frac{1}{n(n-1)(n-2)} - \cdots + (-1)^{n-1}\frac{1}{n!}$$

$$= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1}\frac{1}{n!}$$

这个问题若直接计算有利场合数，无疑是十分复杂的．

四、可列可加性与连续性

下面我们对可列可加性作进一步讨论．从性质 2 知道，由可

列可加性可以推出有限可加性；但是一般来讲，由有限可加性并不能推出可列可加性．

事实上，若 $A_i \in \mathscr{F}$，$i=1, 2, \cdots$且两两互不相容，则由概率的有限可加性只能推出（1.5.2）成立，即

$$P\Big(\sum_{i=1}^{n} A_i\Big) = \sum_{i=1}^{n} P(A_i) \tag{1.5.10}$$

这个等式的左边对任意 n 都不超过 1，因此右边的正项级数收敛．这样应有

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{n} A_i\Big) = \lim_{n\to\infty}\sum_{i=1}^{n} P(A_i) = \sum_{i=1}^{\infty} P(A_i) \tag{1.5.11}$$

与（1.5.1）比较一下就可以知道，为了具有可列可加性，还需要下式成立：

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{n} A_i\Big) = P\Big(\sum_{i=1}^{\infty} A_i\Big) \tag{1.5.12}$$

或者写成更富有启发性的等式：

$$\lim_{n\to\infty} P\Big(\sum_{i=1}^{n} A_i\Big) = P\Big(\lim_{n\to\infty}\sum_{i=1}^{n} A_i\Big) \tag{1.5.13}$$

即要允许把极限号移到概率号里面去，这就提出了一个新要求．

现在就来考察这个新要求．若记

$$S_n = \sum_{i=1}^{n} A_i$$

则 $S_n \in \mathscr{F}$，$n=1, 2, \cdots$而且 $S_n \subset S_{n+1}$，即 S_n 是 $\mathscr{F}$ 中一个单调不减的集序列，这时可改写（1.5.13）成下式：

$$\lim_{n\to\infty} P(S_n) = P(\lim_{n\to\infty} S_n) \tag{1.5.14}$$

一般，对于 $\mathscr{F}$ 上的集合函数，若它对 $\mathscr{F}$ 中任何一个单调不减的集序列 $\{S_n\}$ 均成立（1.5.14），则我们称它是**下连续的**．因此前面的推导表明，为保证概率的可列可加性成立，除要求它具有有限可加性外，还要求它是下连续的．

下面的定理明确地阐明了这三个概念之间的关系．

定理 1.5.1　若 P 是 $\mathscr{F}$ 上满足 $P(\Omega)=1$ 的非负集合函数，

则它具有可列可加性的充要条件为

(i) 它是有限可加的；

(ii) 它是下连续的．

[证明] 充分性只要重新考察前面的推导过程．实际上，若沿用有关记号，则有限可加性保证了（1.5.10）的成立，而下连续性保证了（1.5.13）即（1.5.12）的成立，通过（1.5.11）立刻得到可列可加性．

必要性：其中（i）我们早已建立．为证（ii），设 $\{S_n\}$ 是 $\mathscr{F}$ 中一个单调不减的集序列，那么

$$\bigcup_{i=1}^{\infty} S_i = \lim_{n\to\infty} S_n$$

若定义 $S_0=\varnothing$，则

$$\bigcup_{i=1}^{\infty} S_i = \sum_{i=1}^{\infty}(S_i - S_{i-1})$$

这里的 (S_i-S_{i-1})，$i=1, 2, \cdots$由于 S_i 的单调性，显然两两互不相容，因此由可列可加性得

$$P\Big(\bigcup_{i=1}^{\infty} S_i\Big) = \sum_{i=1}^{\infty} P(S_i - S_{i-1}) = \lim_{n\to\infty}\sum_{i=1}^{n} P(S_i - S_{i-1})$$

但是

$$\sum_{i=1}^{n} P(S_i - S_{i-1}) = P\Big(\sum_{i=1}^{n}(S_i - S_{i-1})\Big) = P(S_n)$$

因此

$$P(\lim_{n\to\infty} S_n) = \lim_{n\to\infty} P(S_n)$$

这就证得了 P 的下连续性．

系 1 概率是下连续的．

系 2 概率是上连续的，即若 $B_i\in\mathscr{F}$，而且 $B_i\supset B_{i+1}$，$i=1, 2, \cdots$则

$$\lim_{n\to\infty} P(B_n) = P(\lim_{n\to\infty} B_n) \tag{1.5.15}$$

[证明] 记 $S_i=\overline{B}_i$，则 $\{S_i\}$ 是单调不减的，由系 1 可得

$$\lim_{n\to\infty}P(S_n)=P(\lim_{n\to\infty}S_n)=P\left(\bigcup_{i=1}^{\infty}S_i\right)=P\left(\overline{\bigcap_{i=1}^{\infty}B_i}\right)$$

因此由（1.5.3），可知

$$1-\lim_{n\to\infty}P(B_n)=1-P\left(\bigcap_{i=1}^{\infty}B_i\right)$$

注意到 $\lim_{n\to\infty}B_n=\bigcap_{i=1}^{\infty}B_i$，即得（1.5.15）

五、概率空间

在科尔莫戈罗夫的概率论公理化结构中，称三元总体$(\Omega,\mathscr{F},P)$为**概率空间**，其中 Ω 是样本空间，$\mathscr{F}$ 是事件域，P 是概率，它们都认为是预先给定的，并以此作为出发点讨论种种问题．至于实际问题中，如何选定 Ω，怎样构造 $\mathscr{F}$，怎样给定 P，则要视具体情况而定．

下面讨论几个具体例子．

［例 7］（有限概率空间）　这时 Ω 中只有 n 个点，这类概率空间在§2 中已有讨论，古典概型即为其特例．在这种场合，一般可取 $\mathscr{F}$ 为 Ω 的所有子集全体，这仍是一个有限的集合，元素总数为 2^n，它满足事件域的三个要求，而且样本点（看作一个单点集）是事件．至于概率，只要对样本点 ω_i 给定满足

$$P(\omega_i)\geqslant 0,\quad i=1,2,\cdots,n$$

$$P(\omega_1)+\cdots+P(\omega_n)=1$$

的一组数 $P(\omega_1)$，$P(\omega_2)$，…，$P(\omega_n)$，那么，若 A 是 $\mathscr{F}$ 中元素，包含样本点 ω_{i_1}，…，ω_{i_k}，则由概率的可加性，自然应令

$$P(A)=P(\omega_{i_1})+\cdots+P(\omega_{i_k})$$

这就给定了事件 A 的概率，从而构成了概率空间$(\Omega,\mathscr{F},P)$．

从这个例子中看到下面两点．

(1) 选定了$(\Omega,\mathscr{F})$之后，对于事件概率的给定还有相当大的灵活性，这表现在 $P(\omega_i)$ 的选取上．因为只有这样，才能用概率空间来描述不同的随机现象．例如在投一次硬币的试验中，Ω

总是由出正面（ω_1）及出反面（ω_2）两个样本点构成．对于均匀的硬币，可以假定它出正面及反面的概率均为$\frac{1}{2}$；但对于很不均匀的硬币，例如出正面可能性大得多的硬币，则必须给定另外的概率，而这只须适当给定 $P(\omega_1)$就可以了．

（2）一旦 $P(\omega_i)$ 给定后，事件 A 的概率并不能任意给定，即在事件域中，各事件的概率有一定关系，给定概率时必须满足这些关系．

［例 8］（离散概率空间）　Ω 由可列个点构成：$\Omega=\{\omega_1, \omega_2, \cdots\}$，这时 $\mathscr{F}$ 还是可以选为 Ω 的子集全体，它满足事件域的三个条件．这时样本点也是事件．为给定概率，可选可列个非负的数 p_i，$i=1, 2, \cdots$满足

$$\sum_{i=1}^{\infty} p_i = 1$$

作为样本点 ω_i 的概率，而一般事件 $A\in\mathscr{F}$ 的概率，则必须取为它所含的样本点的概率之和．

［例 9］　若 $\Omega=\mathbf{R}^1$，即样本空间由全体实数构成，这时 $\mathscr{F}$ 不能取为 Ω 的一切子集，因为这个集类太大，无法在其上定义概率．这时通常取 $\mathscr{F}$ 为直线上博雷尔点集全体 $\mathscr{B}_1$，这是相当大的一个集类，可以把实际问题中所感兴趣的点集都包括在内．另一方面在博雷尔 σ 域上定义概率相当方便，这只要对左闭右开区间给定概率即可．这些我们将在第三章中作深入讨论．

顺便指出，若 Ω 不是 $\mathbf{R}^1$，而是它的一部分，也可类似处理．譬如 Ω 为一个区间，这时 $\mathscr{F}$ 可取为该区间上的博雷尔点集全体：它们通过直线上博雷尔点集与该区间之交而得到．

［例 10］　若 $\Omega=\mathbf{R}^n$（或 $\mathbf{R}^n$ 的一部分），这时可类似于一维场合取 n 维欧几里得空间中的博雷尔点集全体 $\mathscr{B}_n$ 作为事件域 $\mathscr{F}$，在第三章 §2 中将对这种场合进行深入讨论．

第一章小结

本章中介绍了一类新的现象——随机现象，这是一种普遍存

在的现象．在大量随机现象中存在着统计规律性，概率论便是研究随机现象的数量规律的一门数学学科．

"事件"与"概率"是概率论中最基本的两个概念，我们在公理化结构下严格地定义了这两个概念．

为了使读者清楚地理解事件与概率的直观意义，我们采用由具体到抽象，由简单到复杂，由特殊到一般的方式分别介绍了频率、古典概型、几何概率，并从中归纳出事件与概率的本质特征，为公理化定义作准备，这种讲法基本上与概率概念的历史发展平行．

事件的运算及概率的性质是本章的基本内容，也是学习以后各章的必要基础，务必牢固掌握．

我们较完整地研究了古典概型，并介绍了它在产品抽样检查中的应用．对于古典概型的讨论有助于对概率论基本概念的直观理解，而且在以后讨论更一般的情况时，也常以它为特例加以考虑．古典概型中概率的计算有较高的技巧性，读者应该掌握一些最基本的计算方法．

几何概率是很有启发性的一类问题，不过它的严格表述只有用到第三章的一些概念才能做到．

搞清频率与概率的关系是十分重要的一个课题，今后我们将一再回到这个问题上来．

书中我们着重把现代概率论公理化结构作为一个历史发展过程来描述，有兴趣的读者不妨试着进行相反而相成的另一项工作：用公理化结构来概括古典概型、几何概率等特殊模型．

博雷尔点集是一个重要的概念，在第三章中，它将起关键作用．

习 题 一

1．在某城市中，共发行三种报纸 A，B，C．在这城市的居民中，订阅 A 的占 45%，订阅 B 的占 35%，订阅 C 的占 30%，同时订阅 A 及 B 的占 10%，

同时订阅 A 及 C 的占 8%，同时订阅 B 及 C 的占 5%，同时订阅 A,B,C 的占 3%，试求下列百分率：(1) 只订阅 A 的；(2) 只订阅 A 及 B 的；(3) 只订阅一种报纸的；(4) 正好订阅两种报纸的；(5) 至少订阅一种报纸的；(6) 不订阅任何报纸的．

2. 若 A，B，C 是随机事件，说明下列关系式的概率意义：(1) $ABC=A$；(2) $A\cup B\cup C=A$；(3) $AB\subset C$；(4) $A\subset\overline{BC}$.

3. 试把 $A_1\cup A_2\cup\cdots\cup A_n$ 表示成 n 个两两互不相容事件的和．

4. 在某班学生中任选一个同学，以事件 A 表示选到的是男同学，事件 B 表示选到的人不喜欢唱歌，事件 C 表示选到的人是运动员．(1) 表述 $AB\overline{C}$ 及 $A\overline{B}C$；(2) 什么条件下成立 $ABC=A$；(3) 何时成立 $\overline{C}\subset B$；(4) 何时同时成立 $A=B$ 及 $\overline{A}=C$.

5. 用摸球模型造一例，指出样本空间及各种事件运算．

6. 若 A,B,C,D 是四个事件，试用这四个事件表示下列各事件：(1) 这四个事件至少发生一个；(2) A，B 都发生而 C，D 都不发生；(3) 这四个事件恰好发生两个；(4) 这四个事件都不发生；(5) 这四个事件中至多发生一个．

*7. 从 0，1，2，…，9 中随机地取出 5 个数（可重复），以 E_i 记某些数正好出现 i 次这一事件（例如 52353，既属于 E_1，也属于 E_2 及 E_0)，试用文图表示 E_0，E_1，…，E_6 的关系．

8. 证明下列等式：

(1) $\binom{n}{1}+2\binom{n}{2}+3\binom{n}{3}+\cdots+n\binom{n}{n}=n2^{n-1}$

(2) $\binom{n}{1}-2\binom{n}{2}+3\binom{n}{3}+\cdots+(-1)^{n-1}n\binom{n}{n}=0$

(3) $\sum_{k=0}^{a-r}\binom{a}{k+r}\binom{b}{k}=\binom{a+b}{a-r}$

9. 袋中有白球5只，黑球6只，陆续取出三球，求顺序为黑白黑的概率．

10. 一部五卷的文集，按任意次序放到书架上去，试求下列概率：(1) 第一卷出现在旁边；(2) 第一卷及第五卷出现在旁边；(3) 第一卷或第五卷出现在旁边；(4) 第一卷及第五卷都不出现在旁边；(5) 第三卷正好在正中．

11. 把 1，2，3，4，5 诸数各写在一小纸片上，任取其三而排成自左向

右的次序，求所得数是偶数的概率．

12. 在一个装有 n 只白球，n 只黑球，n 只红球的袋中，任取 m 只球，求其中白、黑、红球分别有 m_1，m_2，m_3（$m_1+m_2+m_3=m$）只的概率．

13. 甲袋中有 3 只白球，7 只红球，15 只黑球，乙袋中有 10 只白球，6 只红球，9 只黑球，现从两袋中各取一球，求两球颜色相同的概率．

14. 由盛有号码 1，2，…，N 的球的箱子中有放回地摸了 n 次球，依次记下其号码，试求这些号码按严格上升次序排列的概率．

15. 在上题中这些号码按上升（不一定严格）次序排列的概率．

16. 任意从数列 1，2，…，N 中不放回地取出 n 个数并按大小排列成：$x_1<x_2<\cdots<x_m<\cdots<x_n$，试求 $x_m=M$ 的概率．这里 $1\leqslant M\leqslant N$.

*17. 上题中，若采用有放回取数，这时 $x_1\leqslant x_2\leqslant\cdots\leqslant x_m\leqslant\cdots\leqslant x_n$，试求 $x_m=M$ 的概率．

18. 从 6 双不同的手套中任取 4 只，问其中恰有一双配对的概率是多少？

19. 从 n 双不同的鞋子中任取 $2r$（$2r<n$）只，求下列事件发生的概率：(1) 没有成对的鞋子；(2) 只有一对鞋子；(3) 恰有两对鞋子；(4) 有 r 对鞋子．

20. 袋中有 n 只球，记有号码 1，2，…，n，求下列事件的概率：(1) 任意取出 2 球，号码为 1，2；(2) 任意取出 3 球，没有号码 1；(3) 任意取出 5 球，号码 1，2，3 中至少出现一个．

21. 袋中装有 1，2，…，N 号的球各一只，采用 (1) 有放回；(2) 不放回方式摸球，试求在第 k 次摸球时首次摸到 1 号球的概率．

22. m 个男孩和 n 个女孩（$n\leqslant m$）随机地沿着圆桌坐下，试求任意两个女孩都不相邻的概率．

23. (分赌注问题)　甲、乙二人各出赌注 a，约定谁先胜三局则赢得全部赌注，现已赌三局，甲二胜一负，这时因故中止赌博，若二人赌技相同，问应如何分配赌注，才算公平合理？

24. 从 52 张扑克牌中任意取出 13 张来，问有 5 张黑桃，3 张红心，3 张方块，2 张草花的概率．

25. 桥牌游戏中（四人各从 52 张纸牌中分得 13 张），求 4 张 A 集中在一个人手中的概率．

*26. 在扑克牌游戏中（从 52 张牌中任取 5 张），求下列事件的概率：

(1) 以 A 打头的同花顺次五张牌；(2) 其他同花顺次五张牌；(3) 有四张牌同点数；(4) 三张同点数且另两张也同点数；(5) 五张同花；(6) 异花顺次五张牌；(7) 三张同点数，另外两张不同点数；(8) 五张中有两对；(9) 五张中有一对；(10) 其他情况．

27. 某码头只能容纳一只船，现预知某日将独立来到两只船，且在 24 小时内各时刻来到的可能性都相等，如果它们需要停靠的时间分别为 3 小时及 4 小时，试求有一船要在江中等待的概率．

28. 两人约定于 7 点到 8 点在某地会面，试求一人要等另一人半小时以上的概率．

29. 在一线段 AB 中随机地取两个点把线段截为三段，求这三段可以构成一个三角形的概率(三线段能构成三角形的充要条件是任意二边之和大于第三边)．

30. 在线段 [0，1] 上任意投三个点，问由 0 至三点的三线段能构成三角形与不能构成三角形这两个事件中哪一个事件的概率大．

31. 在一张打上方格的纸上投一枚直径为 1 的硬币，方格要多小才能使硬币与线不相交的概率小于 1%.

32. 从 (0，1) 中随机地取两个数，求下列概率：(1) 两数之和小于 1.2；(2) 两数之积小于 $\frac{1}{4}$；(3) 以上两个要求同时满足．

33. 设 A_1，A_2，…，A_n 是随机事件，试用归纳法证明下列公式

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j) + \sum_{1 \leqslant i < j < k \leqslant n} P(A_i A_j A_k) - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n)$$

34. 某班有 N 个士兵，每人各有一支枪，这些枪外形完全一样，在一次夜间紧急集合中，若每人随机地取走一支枪，问至少有一个人拿到自己的枪的概率．

*35. 在上题中求恰好有 k $(0 \leqslant k \leqslant N)$ 个人拿到自己的枪的概率．

*36. 考试时共有 N 张考签，n 个学生参加考试 $(n \geqslant N)$，被抽过的考签立刻放回，求在考试结束之后，至少有一张考签没有被抽到的概率．

*37. 甲，乙，丙三人按下面规则进行比赛，第一局由甲，乙参加而丙轮空，由第一局的优胜者与丙进行第二局比赛，而失败者则轮空，比赛用这种

方式一直进行到其中一个人连胜两局为止，连胜两局者成为整场比赛的优胜者，若甲，乙，丙胜每局的概率各为$\frac{1}{2}$，问甲，乙，丙成为整场比赛优胜者的概率各是多少？

*38. 父，母，子三人举行比赛，每局总有一人胜一人负（没有和局），每局的优胜者就与未参加此局的人再进行比赛，如果某人首先胜了两局，则他就是整个比赛的优胜者，由父决定第一局由哪两人参加，其中儿子实力最强，所以父为了使自己得胜的概率达到最大，就决定第一局由他与妻子先比赛，试证父的决策为最优策略（任何一对选手中一人胜对方的概率在整个比赛中是不变的）．

39. 给定 $p = P(A), q = P(B), r = P(A \cup B)$，求 $P(A\overline{B})$ 及 $P(\overline{A}\,\overline{B})$．

*40. 已知：$P(AB) = P(A)P(B), C \supset AB, \overline{C} \supset \overline{A}\,\overline{B}$，证明：$P(AC) \geqslant P(A)P(C)$．

41. 设 p_1, p_2, p_{12} 是给定的实数，试证存在两个事件 A_1 及 A_2 使得 $P(A_1) = p_1, P(A_2) = p_2, P(A_1A_2) = p_{12}$ 的充要条件是下列四个不等式同时成立：

$$p_{12} \geqslant 0, p_1 - p_{12} \geqslant 0, p_2 - p_{12} \geqslant 0, 1 - p_1 - p_2 + p_{12} \geqslant 0$$

*42. 用概率论想法求 N 阶行列式的展开式中包含主对角线元素的项数．

*43. 利用概率论的想法证明下列恒等式：

$$1 + \frac{A-a}{A-1} + \frac{(A-a)(A-a-1)}{(A-1)(A-2)} + \cdots + \frac{(A-a)\cdots 2 \cdot 1}{(A-1)\cdots(a+1)a} = \frac{A}{a}$$

其中 A，a 都是正整数，且 $A>a$.

*44. 从一只装有 100 只灯泡的箱子中任抽 5 只灯泡，发现有 2 只是次品，你对此批灯泡的次品数作何估计？（比较用最大似然估计法所得结果与用频率估计概率法的结果是否相同．这种抽查当然用不放回方式．）

45. 求包含事件 A，B 的最小 σ 域．

46. 证明：Ω 的一切子集组成的集类是一个 σ 域．

47. 证明：σ 域之交仍为 σ 域．

48. 证明：包含一切形为 $(-\infty, x)$ 的区间的最小 σ 域是一维博雷尔 σ 域．

49. 设 Q 是定义在 σ 域上的非负广义实值函数（即可以取有限或无限值的函数），如果它具有可列可加性，并且 $Q(\varnothing) = 0$，则称 Q 为测度，试说

明测度概念是算术中计数概念及几何中长度、面积、体积等概念的推广.

50. 用测度概念解释古典概型、几何概率及概率论公理化结构中关于概率的定义.

第二章　条件概率与统计独立性

§1. 条件概率，全概率公式，贝叶斯公式

一、条件概率

对概率的讨论总是在一组固定的条件限制下进行的．以前的讨论总是假定除此之外再无别的信息可供使用，可是，有时我们却会遇到这样的情况，即已知某一事件 B 已经发生，要求另一事件 A 发生的概率．例如考虑有两个孩子的家庭，假定男女出生率一样，则两个孩子（依大小排列）的性别为（男，男），（男，女），（女，男），（女，女）的可能性是一样的．若以 A 记随机选取的这样一个家庭中有一男孩、一女孩这一事件，则显然 $P(A)=\frac{1}{2}$，但是如果我们预先知道这家庭至少有一个女孩，那么，上述事件的概率便应是$\frac{2}{3}$．

两种情况下算出的概率不同．这也很容易理解，因为在第二种情况下，我们多知道了一个条件：事件 B（这一家庭至少有一女孩）发生，因此我们算得的概率事实上是“在已知事件 B 发生的条件下，事件 A 发生的概率”，这个概率我们将记之为 $P(A|B)$．

这种带有条件的概率很重要，下面我们就来研究它．在给出严格定义之前，先考察一些特殊的场合．

就从上述例子出发．这是一个古典概型问题，样本点总数 $n=4$，有利于事件 A 的场合数 $m_A=2$，因此 $P(A)=\frac{1}{2}$；但是假如已知事件 B 发生，即至少有一女孩，那么可能发生的样本点是（男，

女)，(女，男)，(女，女)，总数为 $m_B=3$，而有利场合（至少有一女孩，而且有一男孩、一女孩）数 $m_{AB}=2$，因此

$$P(A|B)=\frac{2}{3}=\frac{m_{AB}}{m_B}=\frac{\frac{m_{AB}}{n}}{\frac{m_B}{n}}=\frac{P(AB)}{P(B)} \qquad (2.1.1)$$

这式子很重要，虽然我们以特例形式引入，但读者不难证明，它对一般古典概型问题也成立.

在几何概率中，若以 $m(A)$，$m(B)$，$m(AB)$，$m(\Omega)$ 分别记事件 A、B，AB，Ω 所对应点集的测度，且 $m(B)>0$，则

$$P(A|B)=\frac{m(AB)}{m(B)}=\frac{\frac{m(AB)}{m(\Omega)}}{\frac{m(B)}{m(\Omega)}}=\frac{P(AB)}{P(B)}$$

结果与古典概型相同.

对频率也有类似结果，请读者自行验证.

在一般场合，我们将把这个算式作为条件概率的定义.

定义 2.1.1 设 $(\Omega, \mathscr{F}, P)$ 是一个概率空间，$B\in\mathscr{F}$，而且 $P(B)>0$，则对任意 $A\in\mathscr{F}$，记

$$P(A|B)=\frac{P(AB)}{P(B)} \qquad (2.1.2)$$

并称 $P(A|B)$ 为**在事件 B 发生的条件下事件 A 发生的条件概率（conditional probability）**.

若未经特别指出，今后出现条件概率 $P(A|B)$ 时，都假定 $P(B)>0$.

由 (2.1.2) 立刻得到

$$P(AB)=P(B)P(A|B) \qquad (2.1.3)$$

这个等式被称为概率的**乘法公式**或**乘法定理**.

若还有 $P(A)>0$，则也可定义 $P(B|A)$，这时有

$$P(AB)=P(A)P(B|A)=P(B)P(A|B) \qquad (2.1.4)$$

[例 1] 在肝癌普查中发现，某地区的自然人群中，每十万

人内平均有 40 人患原发性肝癌，有 34 人甲胎球蛋白高含量，有 32 人既患原发性肝癌又出现甲胎球蛋白高含量.

从这个地区的居民中任抽一人，若他患有原发性肝癌则记为 C，甲胎球蛋白高含量记为 D，这时

$$P(C)=0.0004, P(D)=0.00034, P(CD)=0.00032$$

由条件概率定义可得

$$P(D|C)=\frac{P(CD)}{P(C)}=\frac{0.00032}{0.0004}=0.8$$

$$P(C|D)=\frac{P(CD)}{P(D)}=\frac{0.00032}{0.00034}=0.9412$$

通过计算得知，患原发性肝癌的人有 80%其甲胎球蛋白呈现出高含量，而甲胎球蛋白的测定大大有助于发现原发性肝癌患者：若出现高含量，则有高达 94%以上的概率对患原发性肝癌作出正确诊断.

由于事件 D 的发生，使事件 C 发生的概率由 0.0004 一下子上升到 0.9412. 可见，事件发生的概率，与条件有关，也即与信息有关.

下面讨论条件概率的性质.

首先，不难验证条件概率 $P(A|B)$ 具有概率的三个基本性质：非负性、规范性、完全可加性.

(i) $P(A|B)\geqslant 0$；

(ii) $P(\Omega|B)=1$；

(iii) $P\left(\sum_{i=1}^{\infty}A_i|B\right)=\sum_{i=1}^{\infty}P(A_i|B)$.

因此，类似于概率，对条件概率也可由三个基本性质导出其它一些性质，例如

$$P(\varnothing|B)=0$$

$$P(A|B)=1-P(\bar{A}|B)$$

$$P(A_1\cup A_2|B)=P(A_1|B)+P(A_2|B)-P(A_1A_2|B)$$

特别当 $B=\Omega$ 时，条件概率化为无条件概率，因此把一般的概率看作条件概率也未尝不可.

可以把乘法公式推广到任意 n 个事件之交的场合：

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1}) \tag{2.1.5}$$

这里当然要求 $P(A_1A_2\cdots A_{n-1})>0$.

［例 2］（波利亚罐模型）罐中有 b 只黑球及 r 只红球，随机取出一只，把原球放回，并加进与抽出球同色的球 c 只，再摸第二次，这样下去共摸了 n 次，问前面的 n_1 次出现黑球，后面的 $n_2=n-n_1$ 次出现红球的概率是多少？

［解］ 以 A_1 表示第一次摸出黑球这一事件，……A_{n_1} 表示第 n_1 次摸出黑球，A_{n_1+1} 表示第 n_1+1 次摸出红球，……A_n 表示第 n 次摸出红球. 则

$$P(A_1)=\frac{b}{b+r},\ P(A_2|A_1)=\frac{b+c}{b+r+c}$$

$$P(A_3|A_1A_2)=\frac{b+2c}{b+r+2c},\ \cdots$$

$$P(A_{n_1}|A_1\cdots A_{n_1-1})=\frac{b+(n_1-1)c}{b+r+(n_1-1)c}$$

$$P(A_{n_1+1}|A_1\cdots A_{n_1})=\frac{r}{b+r+n_1c}$$

$$P(A_{n_1+2}|A_1\cdots A_{n_1+1})=\frac{r+c}{b+r+(n_1+1)c},\cdots$$

$$P(A_n|A_1\cdots A_{n-1})=\frac{r+(n_2-1)c}{b+r+(n-1)c}$$

因此

$$P(A_1A_2\cdots A_n)=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}\cdot\frac{b+2c}{b+r+2c}\cdots\frac{b+(n_1-1)c}{b+r+(n_1-1)c}\cdot\frac{r}{b+r+n_1c}\cdot\frac{r+c}{b+r+(n_1+1)c}$$

$$\cdots\frac{r+(n_2-1)c}{b+r+(n-1)c}$$

注意这个答案只与黑球及红球出现次数有关，而与出现的顺序无关.

这个模型曾被波利亚（Pólya）用来作为描述传染病的数学模型. 这是很一般的摸球模型，特别取 $c=0$，则是有放回摸球；取 $c=-1$，则是不放回摸球.

二、全概率公式

概率论的重要研究课题之一是希望从已知的简单事件的概率推算出未知的复杂的事件的概率. 为达到这个目的，经常把一个复杂事件分解为若干个不相容的简单事件之和，再通过分别计算这些简单事件的概率，最后利用概率的可加性得到最终结果. 这里，全概率公式起着很重要的作用.

设事件 A_1，A_2，…，A_n，…是样本空间 Ω 的一个分割，即 A_i（$i=1$，2，…，n，…）两两互不相容，而且

$$\sum_{i=1}^{\infty}A_i=\Omega$$

这样一来

$$B=\sum_{i=1}^{\infty}A_iB$$

这里的 A_iB（$i=1$，2，…，n，…）也两两互不相容（参看图 8）.

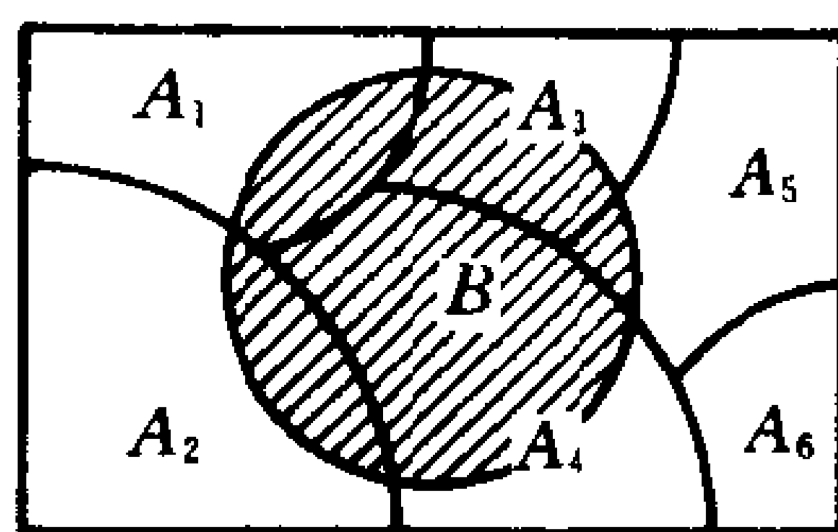


图 8 样本空间的分割

由概率的完全可加性

$$P(B)=\sum_{i=1}^{\infty}P(A_iB)$$

再利用乘法公式即得

$$P(B)=\sum_{i=1}^{\infty}P(A_i)P(B|A_i) \qquad (2.1.6)$$

这公式称为**全概率公式**，它是概率论的一个基本公式，有着多方

面的应用.

从推导中可以看出，当 $P(A_i)=0$ 时，只要把相应的项当作 0 即可. 在多数问题中，(2.1.6) 式只含有限项.

[例 3] 送检的两批灯管在运输中各打碎一支. 若每批十支，而第一批中有 1 支次品，第二批中有 2 支次品. 现从剩下的灯管中任取一支，问抽得次品的概率是多少?

[解] 以 B 表示抽得次品这一事件，为求其概率，有下述两种解题思路，都可以得到最终答案.

思路一. 考虑抽得的灯管出自哪一批，以 A 表示灯管出自第一批，这样一来，$\overline{A}$ 表示灯管出自第二批. 显然 $P(A)=\frac{9}{18}$，$P(\overline{A})=\frac{9}{18}$. 为计算 $P(B|A)$，必须分别考虑第一批中打碎的是次品还是好品，从而得到 $P(B|A)=\frac{1}{10}\times 0+\frac{9}{10}\times\frac{1}{9}=\frac{1}{10}$，同理 $P(B|\overline{A})=\frac{2}{10}\times\frac{1}{9}+\frac{8}{10}\times\frac{2}{9}=\frac{2}{10}$，因此

$$P(B)=P(A)P(B|A)+P(\overline{A})P(B|\overline{A})=\frac{3}{20}$$

思路二. 直接分析打碎灯管的各种可能情况，以事件 A_1，A_2，A_3，A_4 分别记前后两批中打碎者为 (次，次)，(次，好)，(好，次)，(好，好)，其概率为 $P(A_1)=\frac{2}{100}$，$P(A_2)=\frac{8}{100}$，$P(A_3)=\frac{18}{100}$，$P(A_4)=\frac{72}{100}$. 在事件 A_i 发生的条件下，事件 B 的条件概率容易求得为 $P(B|A_1)=\frac{1}{18}$，$P(B|A_2)=\frac{2}{18}$，$P(B|A_3)=\frac{2}{18}$，$P(B|A_4)=\frac{3}{18}$，因此利用全概率公式，

$$P(B)=\sum_{i=1}^{4}P(A_i)P(B|A_i)=\frac{3}{20}$$

两种解法都是利用全概率公式. 从例中还可看出全概率公式之所以有力，就在于它概括了一种普遍的解题思路：化整为零，各

个击破.

三、贝叶斯（Bayes）公式

若事件 B 能且只能与两两互不相容的事件 A_1，A_2，…，A_n，…之一同时发生，即

$$B=\sum_{i=1}^{\infty}BA_i$$

由于

$$P(A_iB)=P(B)P(A_i|B)=P(A_i)P(B|A_i)$$

故

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{P(B)}$$

再利用全概率公式即得

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_{i=1}^{\infty}P(A_i)P(B|A_i)} \qquad (2.1.7)$$

这个公式称为**贝叶斯公式**.

贝叶斯公式在概率论和数理统计中有着多方面的应用. 假定 A_1、A_2，…是导致试验结果的“原因”，$P(A_i)$ 称为**先验概率**，它反映了各种“原因”发生的可能性大小，一般是以往经验的总结，在这次试验前已经知道. 现在若试验产生了事件 B，这个信息将有助于探讨事件发生的“原因”. 条件概率 $P(A_i|B)$ 称为**后验概率**，它反映了试验之后对各种“原因”发生的可能性大小的新知识. 例如在医疗诊断中，有人为了诊断病人到底是患了疾病 A_1，A_2，…，A_n 中的哪一种，对病人进行观察与检查，确定了某个指标 B（譬如是体温、脉搏、血液中转氨酶含量等等），他想用这类指标来帮助诊断. 这时就可以用贝叶斯公式来计算有关概率. 首先必须确定先验概率 $P(A_i)$，这实际上是确定人患各种疾病的可能性大小，以往的资料可以给出一些初步数据；其次是要确定 $P(B|A_i)$，这里当然主要依靠医学知识. 有了它们，利用贝叶斯公式就可算出

$P(A_i|B)$. 显然，对应于较大 $P(A_i|B)$ 的"病因" A_i，应多加考虑. 在实际工作中，检查的指标 B 一般有多个，综合所有的后验概率，当然会对诊断有很大帮助. 在实现计算机自动诊断或辅助诊断中，这方法是有实用价值的.

下面介绍应用贝叶斯公式的几个例子.

[例 4] 在数字通信中，由于存在着随机干扰，因此接收到的信号与发出的信号可能不同，为了确定发出的信号，通常要计算各种概率，下面只讨论一种比较简单的模型——二进位信道.

若发报机以 0.7 和 0.3 的概率发出信号 0 和 1（譬如分别用低电平与高电平表示），由于随机干扰的影响，当发出信号 0 时，接收机不一定收到 0，而是以概率 0.8 和 0.2 收到信号 0 和 1；同样地，当发报机发出信号 1 时，接收机以概率 0.9 和 0.1 收到信号 1 和 0. 其关系如图 9 所示.

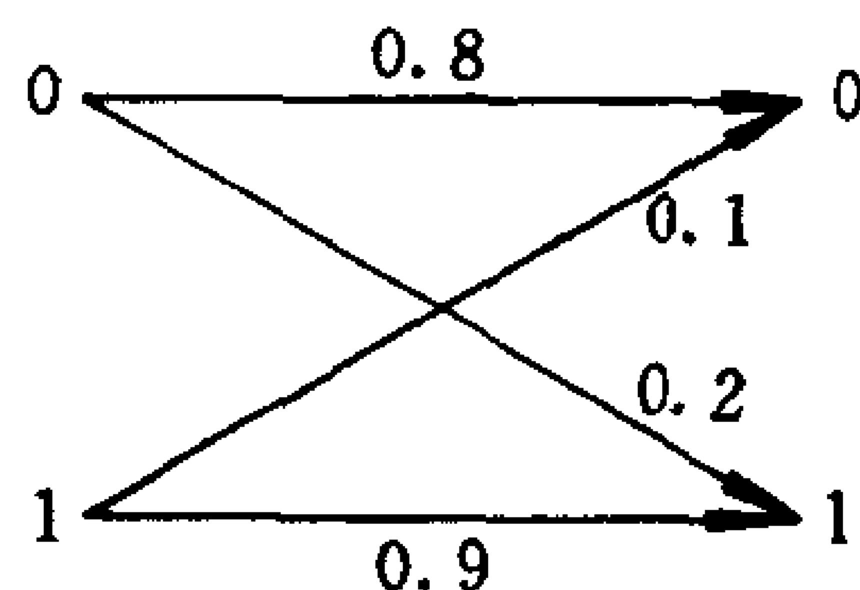


图 9

假如接收机收到信号 0，则这时有两种可能情况：第一种情况是发报机确实发出信号 0，而信号传输中没有出错；第二种情况是发报机发出信号 1，但是传输中出现错误，因此接收到信号 0. 所以在这类问题中，计算"当接收机收到信号 0 时，发报机是发出信号 0 的概率"便很有必要.

把发报机发出信号 0 记为事件 A_0，发出信号 1 记为事件 A_1，接收机接到信号 0 记为事件 B，我们要求的是 $P(A_0|B)$.

由于 $P(A_0)=0.7, P(A_1)=0.3, P(B|A_0)=0.8, P(B|A_1)=0.1$，用贝叶斯公式，

$$P(A_0|B)=\frac{P(A_0)P(B|A_0)}{P(A_0)P(B|A_0)+P(A_1)P(B|A_1)}$$
$$=\frac{0.7\times 0.8}{0.7\times 0.8+0.3\times 0.1}=\frac{0.56}{0.59}=0.949$$

[例 5] 假定用血清甲胎球蛋白法诊断肝癌，$P(A|C)=$

0.95，$P(A|\bar{C})=0.90$，这里 C 表示被检验者患有肝癌这一事件，A 表示判断被检验者患有肝癌这一事件．又设在自然人群中 $P(C)=0.0004$．现在若有一人被此检验法诊断为患有肝癌，求此人真正患有肝癌的概率 $P(C|A)$．

［解］　由贝叶斯公式，

$$P(C|A)=\frac{P(C)P(A|C)}{P(C)P(A|C)+P(\bar{C})P(A|\bar{C})}$$

$$=\frac{0.0004\times0.95}{0.0004\times0.95+0.9996\times0.1}=0.0038$$

既然检验法相当可靠，那么为什么用该法诊断为肝癌的人真正患有肝癌的可能性却如此之小呢？经过分析可以看出，主要是先验概率 $P(C)$ 很小．对自然人群来讲肝癌毕竟是一种罕见病．如果我们检查的是一个肝癌的可疑人群，譬如一批普查中筛选出的甲胎球蛋白高含量者，这时相应的先验概率就高得多，例如就取例 1 中的 0.9412，那么

$$P(C|A)=\frac{0.9412\times0.95}{0.9412\times0.95+0.0588\times0.1}=0.9935$$

因此相应的后验概率也就大为提高．实际普查中正是如此一级一级筛选的．

这个数值例子给我们的最大启示是：后验概率的大小很受先验概率选取的影响．

在贝叶斯公式的使用中，最有争议之点就是先验概率的选取．我们上面所举的两个例子中，这些先验概率都是通过以往大量实际调查而得出的，符合概率的频率解释，因此使用中不致于发生疑问．

不过，在贝叶斯公式的使用中也还存在着另一种情况，就是先验概率是由某一种主观的方式给定的，譬如对于未来宏观经济形势的看法，对物价、利率、汇率变化的估计，对某种新型产品上市后受欢迎程度的预估，甚至对某星球上存在生命现象的估计等等．这种把概率解释为信任程度的做法含有明显的主观性，称

为主观概率.

主观概率与贝叶斯学派的发展息息相关，后者在二战后得到很快发展，理论上与决策理论关系密切，并且找到不少应用. 因此对于主观概率的争论看来将长期持续下去.

贝叶斯方法的大量应用出现在决策问题中，下面是自动识别系统的一个例子.

[例 6]　(贝叶斯决策)　为了判断一个字母是"C"还是"O"，通常采用先抽取它的某一个特征 X，然后再根据这个特征作出判决，这时贝叶斯决策是常用方法之一.

以 A_1, A_2 分别记被检验的字母为 C 或 O 这一事件，它们的先验概率 $P(A_1)$ 及 $P(A_2)$ 应预先给定，此外要通过试验确定 $P(X|A_1)$ 及 $P(X|A_2)$，由贝叶斯公式得

$$P(A_i|X) = \frac{P(A_i)P(X|A_i)}{\sum_{i=1}^{2} P(A_i)P(X|A_i)}$$

其中，$i=1, 2$. 若 $P(A_1|X)>P(A_2|X)$，则作出决策：具有特征 X 的字母是 C.

这个方法在模式识别这一新兴学科中有重要应用，当然，这里是大为简化了的模型.

§ 2. 事件独立性

一、两个事件的独立性

本节中我们引进一个新的概念——统计独立性. 先从两个事件的独立性开始，然后讨论更为一般的场合.

还是从考虑古典概型的一个例子作为出发点.

[例 1]　一口袋中装有 a 只黑球和 b 只白球，采用有放回摸球，求：

(1) 在已知第一次摸得黑球的条件下，第二次摸出黑球的概

率；

(2) 第二次摸出黑球的概率.

［解］ 以事件 A 表示第一次摸得黑球，事件 B 表示第二次摸得黑球. 则

$$P(A)=\frac{a}{a+b},P(AB)=\frac{a^2}{(a+b)^2},P(\bar{A}B)=\frac{ba}{(a+b)^2}$$

所以

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{a}{a+b}$$

而

$$P(B)=P(AB)+P(\bar{A}B)=\frac{a^2}{(a+b)^2}+\frac{ba}{(a+b)^2}=\frac{a}{a+b}$$

注意这里的 $P(B|A)=P(B)$，即事件 A 发生与否，对事件 B 发生的概率没有影响. 从直观上讲，这很自然. 因为我们这里采用的是有放回摸球，因此第二次摸球时袋中球的组成与第一次摸球时完全相同，当然第一次摸球的结果实际上不影响第二次摸球. 在这种场合可以说，事件 A 与事件 B 的出现有某种“独立性”.

对此，我们引进

定义 2.2.1 对事件 A 及 B，若

$$P(AB)=P(A)P(B) \tag{2.2.1}$$

则称它们是**统计独立的**，简称**独立的** (independent).

注意，按照这个定义，必然事件 Ω 及不可能事件 $\varnothing$ 与任何事件独立. 此外，从 (2.2.1) 中看出，A 与 B 的位置对称，因此亦称 **A 与 B 相互独立**.

推论 1 若事件 A，B 独立，且 $P(B)>0$，则

$$P(A|B)=P(A) \tag{2.2.2}$$

［证明］ 由条件概率定义及 (2.2.1) 得

$$P(A|B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B)}{P(B)}=P(A)$$

因此，若事件 A，B 相互独立，则 A 关于 B 的条件概率等于无条件概率 $P(A)$，这表示 B 的发生对于事件 A 是否发生没有提供任何信息，独立性就是把这种关系从数学上加以严格定义.

推论 2 若事件 A 与 B 独立，则下列各对事件也相互独立：

$$\{\overline{A}, B\}, \{A, \overline{B}\}, \{\overline{A}, \overline{B}\}$$

[证明] 由于

$$\begin{aligned} P(\overline{A}B) &= P(B-AB) = P(B) - P(AB) \\ &= P(B) - P(A)P(B) = P(B)[1-P(A)] \\ &= P(\overline{A})P(B) \end{aligned}$$

所以 $\overline{A}$ 与 B 相互独立，由它立刻推出 $\overline{A}$ 与 $\overline{B}$ 相互独立，由 $\overline{\overline{A}}=A$ 又推出 A，$\overline{B}$ 相互独立.

不放回摸球模型提供了不独立的一个简单的例子.

[例 2] 在前例中，若采用不放回摸球，试求同样那两个事件的概率.

[解] 这时

$$P(A)=\frac{a}{a+b}, P(AB)=\frac{a(a-1)}{(a+b)(a+b-1)}$$

$$P(\overline{A}B)=\frac{ba}{(a+b)(a+b-1)}$$

所以

$$P(B|A)=\frac{P(AB)}{P(A)}=\frac{a-1}{a+b-1}$$

而

$$P(B)=P(AB)+P(\overline{A}B)=\frac{a}{a+b}$$

这里 $P(B|A) \neq P(B)$，即事件 B 与事件 A 不是相互独立的. 因为第一次摸得黑球，事实上已使袋中球的组成成份改变了，当然要影响第二次摸得黑球的概率. 而 $P(A)=P(B)$，则再次证明抽签与顺序无关.

二、多个事件的独立性

我们先定义三个事件 A，B，C 的独立性.

定义 2.2.2 对于三个事件 A，B，C，若下列四个等式同时成立，则称它们**相互独立**.

$$\left.\begin{aligned} P(AB)=P(A)P(B) \\ P(BC)=P(B)P(C) \\ P(AC)=P(A)P(C) \end{aligned}\right\} \qquad (2.2.3)$$

$$P(ABC)=P(A)P(B)P(C) \qquad (2.2.4)$$

按两个事件独立性的定义，我们知道若（2.2.3）成立，则 A 与 B，B 与 C，C 与 A 都相互独立，即 A，B，C 两两独立.

读者自然会提出这样一个问题：三个事件 A，B，C 两两独立，能否保证它们相互独立呢？即能否由（2.2.3）推出（2.2.4）？回答是否定的，这从下面简单的例子就可看出.

[例 3]（伯恩斯坦反例） 一个均匀的正四面体，其第一面染成红色，第二面染成白色，第三面染成黑色，而第四面同时染上红，白，黑三种颜色. 现在以 A，B，C 分别记投一次四面体出现红，白，黑颜色朝下的事件，则由于在四面体中有两面有红色，因此

$$P(A)=\frac{1}{2}$$

同理 $P(B)=P(C)=\frac{1}{2}$，容易算出

$$P(AB)=P(BC)=P(AC)=\frac{1}{4}$$

所以（2.2.3）成立，即 A，B，C 两两独立，但是

$$P(ABC)=\frac{1}{4}\neq\frac{1}{8}=P(A)P(B)P(C)$$

因此（2.2.4）不成立，从而 A，B，C 不相互独立.

下面再提供一个例子说明由（2.2.4）不能推出（2.2.3），进

一步说明要 A，B，C 相互独立必须同时要求（2.2.3）及（2.2.4）成立.

［例 4］ 若有一个均匀正八面体，其第 1，2，3，4 面染红色，第 1，2，3，5 面染白色，第 1，6，7，8 面染上黑色，现在以 A，B，C 分别表示投一次正八面体出现红，白，黑的事件，则

$$P(A)=P(B)=P(C)=\frac{4}{8}=\frac{1}{2}$$

$$P(ABC)=\frac{1}{8}=P(A)P(B)P(C)$$

但是

$$P(AB)=\frac{3}{8}\neq\frac{1}{4}=P(A)P(B)$$

现在我们可以定义 n 个事件的独立性.

定义 2.2.3 对 n 个事件 A_1，A_2，…，A_n，若对于所有可能的组合 $1\leqslant i<j<k<\cdots\leqslant n$ 成立着

$$\left.\begin{aligned}&P(A_iA_j)=P(A_i)P(A_j)\\&P(A_iA_jA_k)=P(A_i)P(A_j)P(A_k)\\&\cdots\cdots\\&P(A_1A_2\cdots A_n)=P(A_1)P(A_2)\cdots P(A_n)\end{aligned}\right\}\qquad(2.2.5)$$

则称 A_1，A_2，…，A_n **相互独立**.

这里第一行有 $\binom{n}{2}$ 个式子，第二行有 $\binom{n}{3}$ 个式子等等，因此共应满足

$$\binom{n}{2}+\binom{n}{3}+\cdots+\binom{n}{n}=2^n-n-1$$

个等式. 由三个事件的场合可看出同时满足这些关系式是必须的.

显然若 n 个事件相互独立，则它们中的任何 m（$2\leqslant m<n$）个事件也相互独立. 此外对于多个相互独立的事件也成立着类似于上述推论 1 及推论 2 的结果，读者试自行叙述并验证之.

最后，称无穷多个事件是相互独立的，如果其中任意有限多

个事件都相互独立.

三、事件独立性与概率的计算

从事件独立性的定义立刻能看出：若事件是独立的，则许多概率的计算就可以大为简化. 下面先举两个例子.

[相互独立事件至少发生其一的概率的计算]　若 $A_1, A_2, \cdots, A_n$ 是 n 个相互独立的事件，则由于

$$\overline{A_1 \cup A_2 \cup \cdots \cup A_n} = \overline{A}_1 \overline{A}_2 \cdots \overline{A}_n$$

因此

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= 1 - P(\overline{A}_1 \overline{A}_2 \cdots \overline{A}_n) \\ &= 1 - P(\overline{A}_1) P(\overline{A}_2) \cdots P(\overline{A}_n) \end{aligned} \tag{2.2.6}$$

这个公式比起不独立的场合，要简便得多，它经常被用到.

[例 5]　假若每个人血清中含有肝炎病毒的概率为 0.4%，混合 100 个人的血清，求此血清中含有肝炎病毒的概率.

[解]　以 A_i ($i=1, 2, \cdots, 100$) 记第 i 个人的血清含有肝炎病毒这一事件，可以认为它们相互独立，所求概率为 $P(A_1 \cup \cdots \cup A_{100})$，由 (2.2.6) 得

$$P(A_1 \cup \cdots \cup A_{100}) = 1 - P(\overline{A}_1) \cdots P(\overline{A}_{100}) = 1 - 0.996^{100} \approx 0.33$$

虽然每个人有病毒的概率很小，但是混合后则有很大的概率，在实际工作中，这类效应值得充分重视.

[在可靠性理论中的应用]　对于一个元件，它能正常工作的概率 p，称为它的可靠性. 元件组成系统，系统正常工作的概率称为该系统的可靠性. 随着近代电子技术的迅猛发展，关于元件和系统可靠性的研究已发展成为一门新的学科——可靠性理论.

这里，我们通过一些例子来说明有关的概念.

[例 6]　如果构成系统的每个元件的可靠性均为 r，$0<r<1$，且各元件能否正常工作是相互独立的，试求下面附加通路系统的可靠性 [见图 10 (a)].

［解］ 每条通路要能正常工作，当且仅当该通路上各元件正常工作，故其可靠性为

$$R_C=r^n$$

即通路发生故障的概率为 $1-r^n$. 由于系统是由两通路并联而成的，两通路同时发生故障的概率为 $(1-r^n)^2$，因此附加通路系统的可靠性为

$$R_S=1-(1-r^n)^2=r^n(2-r^n)=R_C(2-R_C)$$

注意到 $R_C<1$，故 $R_S>R_C$，所以附加通路能使系统的可靠性增加.

［例 7］ 在前例条件下，试求下面附加元件系统的可靠性［见图 10（b）］.

［解］ 每对并联元件的可靠性为

$$R'=1-(1-r)^2=r(2-r)$$

系统由各对并联元件串联而成，故其可靠性为

$$R'_S=(R')^n=r^n(2-r)^n$$

显然 $R'_S>R_C$. 因此用附加元件的方法同样也能增加系统的可靠性.

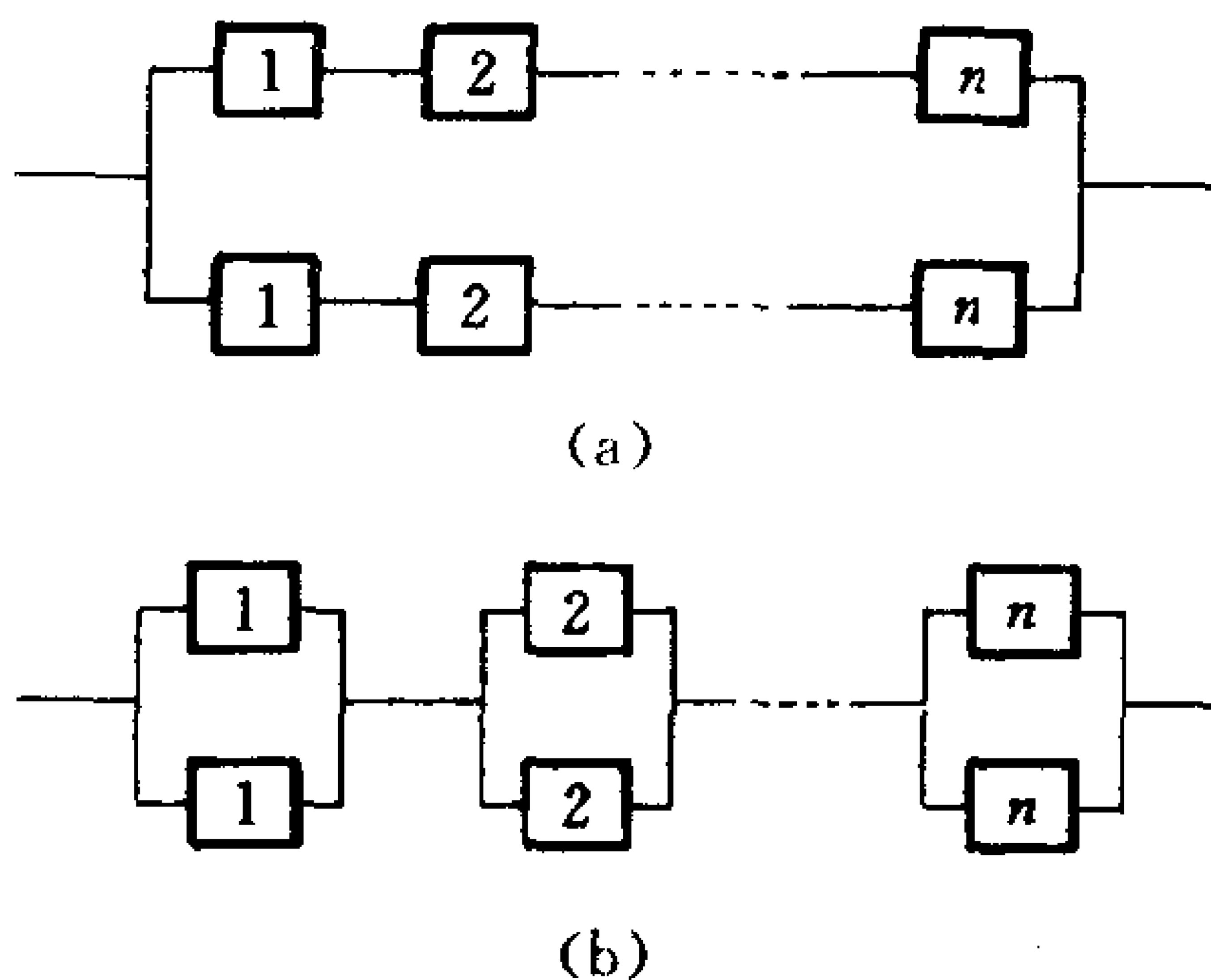


图 10 附加备份系统

利用数学归纳法不难证明当 $n\geqslant 2$ 时，$(2-r)^n>2-r^n$，即 $R'_S>R_S$. 因此虽然上面两个系统同样由 $2n$ 个元件构成，作用也相同，但是第二种构成方式比第一种方式可靠性来得大. 寻找可靠性较大的构成方式也是可靠性理论的研究课题之一.

从上述讨论可以看出，元件与系统的可靠性是用概率来定义的，所以概率论是研究可靠性理论的重要工具.

四、试验的独立性

有了事件独立性的概念，我们可以定义试验的独立性．直观上，试验 E_1 与试验 E_2 要能说是独立的，应是指试验 E_1 的结果的发生与试验 E_2 的结果的发生是独立的，所以自然想到通过各试验的事件间的独立性来定义试验的独立性．为了做到这点，首先要构造一个能描述这些试验的公共的样本空间.

设试验 E_1 的样本空间是 $\Omega_1=\{\omega^{(1)}\}$，试验 E_2 的样本空间是 $\Omega_2=\{\omega^{(2)}\}$，……试验 E_n 的样本空间是 $\Omega_n=\{\omega^{(n)}\}$，为描述这 n 次试验，应构造复合试验 E，它表示依次进行试验 E_1，E_2，…，E_n，其样本点为

$$\omega=(\omega^{(1)},\ \omega^{(2)},\ \cdots,\ \omega^{(n)})$$

这个样本空间记作

$$\Omega=\Omega_1\times\Omega_2\times\cdots\times\Omega_n$$

［例 8］　若试验 E_1 是掷一枚硬币，$\Omega_1=\{$正，反$\}$，试验 E_2 是从装有红白黑三球的袋子中摸出一球，$\Omega_2=\{$红，白，黑$\}$，则复合试验 E 表示先掷一枚硬币再摸一球，它相应的样本空间 $\Omega=\Omega_1\times\Omega_2$ 由下列 6 个样本点构成：（正，红），（正，白），（正，黑），（反，红），（反，白），（反，黑）.

接着我们可以引进“与第 k 次试验有关的事件”的概念．这种事件发生与否仅与第 k 次试验的结果有关，即为了判断某一样本点是否属于这个事件，只须察看它的第 k 个分量．例如在例 8 中，“第 2 次试验摸出红球”这一事件就是一个与第 2 次试验有关的事件．值得指出，必然事件 Ω 及不可能事件 $\varnothing$ 可以认为与所有的试验有关.

现在若以 $\mathscr{A}_k$ 记与第 k 次试验有关的事件全体，则可以通过下列方式定义试验的独立性.

定义 2.2.4 若对于任意的

$$A^{(1)}\in\mathscr{A}_1,\ A^{(2)}\in\mathscr{A}_2,\ \cdots,\ A^{(n)}\in\mathscr{A}_n$$

均成立

$$P(A^{(1)}A^{(2)}\cdots A^{(n)})=P(A^{(1)})P(A^{(2)})\cdots P(A^{(n)})$$

则称试验 E_1, E_2, …, E_n **是相互独立的**.

注意到 $\Omega\in\mathscr{A}_i$, $i=1, 2, \cdots, n$, 因此由定义立刻推出, 若 n 个试验相互独立, 则其中的 m ($2\leqslant m<n$) 个试验也是相互独立的.

在例 8 中, 若对样本空间 Ω 中的 6 个样本点都给定概率 $\frac{1}{6}$, 则不难验证试验 E_1 与试验 E_2 是相互独立的. 读者可以思考一下, 是否还可以给定其他概率使试验 E_1 与试验 E_2 独立?

n 次有放回摸球所构成的 n 个试验是相互独立的, 而 n 次不放回摸球模型则是 n 个试验不独立的简单例子.

特别重要的一类试验是所谓**重复独立试验**, 这时 $\Omega_1=\Omega_2=\cdots=\Omega_n$, 有关事件的概率保持不变, 而且各次试验是相互独立的. 投 n 次硬币或进行 n 次有放回摸球是重复独立试验的简单例子. 重复独立试验是作为"在同样条件下重复试验"的数学模型而出现的, 它在概率论中很有地位, 因为随机现象的统计规律性只有在大量重复试验中才会显示出来.

我们在下一节将研究一类最简单的重复独立试验——伯努利 (Bernoulli) 试验.

§3. 伯努利试验与直线上的随机游动

一、伯努利概型

在许多问题中, 我们对试验感兴趣的是试验中某事件 A 是否发生. 例如在产品抽样检查中注意的是抽到废品, 还是抽到好品; 在掷硬币时注意的是出正面还是出反面. 在这类问题中我们可以

把事件域取为$\mathscr{F}=\{\varnothing, A, \overline{A}, \Omega\}$，并称出现$A$为“成功”，出现$\overline{A}$为“失败”．这种只有两个可能结果的试验称为**伯努利试验**．

有些试验的结果不止两个，譬如，在电报传输中，既要传送字母A，B，…，Z等，又要传送其他符号．但是假如我们所关心的只是字母在传送中所占的百分比，而不再区别到底是哪一个字母，则我们可以把出现字母当作是成功，出现其他符号一律当作是失败，这时就可以把问题看作伯努利试验．同样地，电子管的寿命可以是不小于0的任一数值，但是有时根据需要，我们可以把寿命大于5000小时的电子管当作合格品，其余都作为次品．那么，这类问题还是可以归结为伯努利试验．这种例子可以举出不少．

在伯努利试验中，首先是要给出下面概率

$$P(A)=p, P(\overline{A})=q \tag{2.3.1}$$

显然$p\geqslant 0$，$q\geqslant 0$，且$p+q=1$．

现在考虑重复进行n次独立的伯努利试验，这里的“重复”，是指在每次试验中事件A，从而事件$\overline{A}$出现的概率都保持不变．这种试验称为**n重伯努利试验**，记作E^n．

总之，n重伯努利试验有下面四个约定：

(i) 每次试验至多出现两个可能结果之一：A或$\overline{A}$；

(ii) A在每次试验中出现的概率p保持不变；

(iii) 各次试验相互独立；

(iv) 共进行n次试验．

下面先给出n重伯努利试验的概率空间．

n重伯努利试验E^n的样本点形如：

$$(\omega_1, \omega_2, \cdots, \omega_n) \tag{2.3.2}$$

其中ω_i是A_i或$\overline{A}_i$，分别表示第i次试验中出现A或$\overline{A}$，显然这种样本点共有2^n个，这是一个有限样本空间．

像对所有有限样本空间一样，可以把样本点的任意子集作为事件，构成事件域，不过在这个场合一般不用明显写出．

为了给定概率，只要利用试验的独立性和通过（2.3.1）给定了 p，就能给出形如（2.3.2）的样本点的概率，从而也就给定了所有事件的概率. 例如（A_1，A_2，…，A_{n-1}，$\overline{A}_n$）表示前 $n-1$ 次试验均出现事件 A 而第 n 次试验出现事件 $\overline{A}$，这是一个样本点，其概率为 $p^{n-1}q$. 一般，若一个样本点中，A 出现 k 次，因而 $\overline{A}$ 出现 $n-k$ 次，则其概率均为 p^kq^{n-k}，跟 A 及 $\overline{A}$ 出现的顺序无关.

为书写方便起见，上举样本点将简记为 $A_1A_2\cdots A_{n-1}\overline{A}_n$，其余的可以类推.

有时也需要考虑**可列重伯努利试验** E^∞ 的场合，这时样本点形如

$$(\omega_1,\ \omega_2,\ \cdots,\ \omega_n,\ \cdots) \tag{2.3.3}$$

其中 ω_i 仍取 A_i 或 $\overline{A}_i$，这时样本空间不再有限，甚至也不可列，事实上它可与［0，1］区间进行一一对应，对这种试验就不能把样本空间的任意子集都看作事件.

伯努利试验是一种非常重要的概率模型，它是“在同样条件下进行重复试验”的一种数学模型，特别在讨论某事件出现的频率时常用这种模型. 历史上，伯努利概型是概率论中最早研究的模型之一，也是得到最多研究的模型之一，在理论上具有重要意义；另一方面，它有着广泛的实际应用，例如在工业产品质量检查中，在群体遗传学中它都占有重要地位. 在我们这门课程中，一些较为深入的结果也是结合伯努利概型进行讨论的.

二、伯努利概型中的一些分布

下面我们计算伯努利概型中所出现的一些事件的概率，这些概率非常重要.

1. 伯努利分布

若只进行一次伯努利试验，则或是事件 A 出现，或是事件 $\overline{A}$ 出现，其概率由（2.3.1）给出，称为**伯努利分布**，这是最简单的情况.

2. 二项分布

我们来确定 n 重伯努利试验中事件 A 出现 k 次的概率，这概率我们记之为 $b(k;n,p)$.

若以 B_k 记 n 重伯努利试验中事件 A 正好出现 k 次这一事件，而以 A_i 表示第 i 次试验中出现事件 A，以 $\overline{A}_i$ 表示第 i 次试验中出现 $\overline{A}$，则

$$B_k=A_1A_2\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n+\cdots+\overline{A}_1\overline{A}_2\cdots\overline{A}_{n-k}A_{n-k+1}\cdots A_n \tag{2.3.4}$$

右边的每一项表示在某 k 次试验中出现事件 A，在另外 $n-k$ 次试验中出现 $\overline{A}$. 这种项共有 $\binom{n}{k}$ 个，而且两两互不相容. 由于试验的独立性

$$P(A_1A_2\cdots A_k\overline{A}_{k+1}\cdots\overline{A}_n)=P(A_1)P(A_2)\cdots P(A_k)P(\overline{A}_{k+1})\cdots P(\overline{A}_n)=p^kq^{n-k} \tag{2.3.5}$$

同理可得（2.3.4）中右边各项所对应事件的概率均为 p^kq^{n-k}，利用概率的可加性知

$$P(B_k)=\binom{n}{k}p^kq^{n-k}$$

即

$$b(k;n,p)=\binom{n}{k}p^kq^{n-k},\ k=0,1,2,\cdots,n \tag{2.3.6}$$

注意到 $b(k;n,p)$，$k=0,1,2,\cdots,n$，是二项式 $(q+ps)^n$ 展开式中 s^k 项的系数，因此 (2.3.6) 称为**二项分布** (binomial distribution). 特别地

$$\sum_{k=0}^{n}b(k;n,p)=\sum_{k=0}^{n}\binom{n}{k}p^kq^{n-k}=(q+p)^n=1 \tag{2.3.7}$$

[例 1]　若在 N 件产品中有 M 件废品，现进行 n 次有放回的抽样检查，问共抽得 k 件废品的概率是多少？

[解]　由于抽样是有放回的，因此这是 n 重伯努利试验，若

每次抽查一件

以 A 记各次试验中出现废品这一事件，则

$$p=P(A)=\frac{M}{N}$$

因此所求的概率为

$$b\left(k;\ n,\ \frac{M}{N}\right)=\binom{n}{k}\left(\frac{M}{N}\right)^k\left(1-\frac{M}{N}\right)^{n-k}$$

这概率在第一章§3中曾出现过.

[例 2]　在群体遗传学中，假定可遗传的指标是依赖于基因的. 基因总是成对出现并且具有两种型式 A 及 a 中的一种. 假定每一代具有 $2N$ 个基因，则其中 A 所占的比例数 $\frac{i}{2N}$ 称为基因频率. 基因频率的变化过程与群体的进化情况有着密切的关系.

如果进行的是随机交配，也就是说任何个体有同样的机会和任何其他个体配种，则遗传学中对基因的遗传作下面假定：子代个体是按伯努利概型从上一代每个亲体中取得基因的.

因此若上一代的基因频率为 $\frac{i}{2N}$，则下一代的基因频率为 $\frac{j}{2N}$ 的概率由二项分布给出

$$p_j=\binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{2N-j}$$

在这个例子中基因频率 $\frac{i}{2N}$ 相当于伯努利试验中出现成功的概率 p.

3. 几何分布

现在讨论在伯努利试验中首次成功出现在第 k 次试验的概率，要使首次成功出现在第 k 次试验，必须而且只须在前 $k-1$ 次试验中都出现事件 $\bar{A}$，而第 k 次试验出现 A，因此若沿用推导二项分布时所用的记号 A_i 及 $\bar{A}_i$，则这事件（记为 W_k）可表为

$$W_k=\bar{A}_1\bar{A}_2\cdots\bar{A}_{k-1}A_k \tag{2.3.8}$$

利用事件的独立性，其概率为

$$P(W_k)=P(\bar{A}_1)P(\bar{A}_2)\cdots P(\bar{A}_{k-1})P(A_k)=q^{k-1}p$$

记

$$g(k;\ p)=q^{k-1}p,\ k=1,\ 2,\ \cdots \tag{2.3.9}$$

$g(k;\ p)$ 是几何级数的一般项，因此（2.3.9）称为**几何分布**. 这里有

$$\sum_{k=1}^{\infty} g(k;\ p)=\sum_{k=1}^{\infty} q^{k-1}p=p\frac{1}{1-q}=1 \tag{2.3.10}$$

几何分布给出了等待事件 A 出现共等了 k 次的概率，这类概率在许多问题中出现，我们在第一章 § 3 中曾遇到它. 下面是一个模型化了的例子.

［例 3］ 一个人要开门，他共有 n 把钥匙，其中仅有一把是能开这门的. 他随机地选取一把钥匙开门，即在每次试开时每一把钥匙都以概率 $\frac{1}{n}$ 被使用，这人在第 s 次试开时才首次成功的概率是多少？

［解］ 这是一个伯努利试验，$p=\frac{1}{n}$，由（2.3.9），所求概率为

$$g\left(s;\ \frac{1}{n}\right)=\left(\frac{n-1}{n}\right)^{s-1}\frac{1}{n}$$

应当指出，讨论事件 A 的首次出现事实上牵涉到可列次伯努利试验，因此它的样本空间应取为（2.3.3）所表示的样本点全体. 这个空间不是可列的，所以不能把它的一切子集都作为事件. 在上述讨论中，我们只是把用（2.3.8）表出的 W_k 作为事件，并以（2.3.9）给出其概率，这里的 W_k 事实上也包含了不可列个用（2.3.3）表出的样本点. 当然在讨论这个问题时，我们也可以干脆把 W_k 作为样本点，这时样本空间取为 $\{W_1, W_2, W_3, \cdots\}$，并以（2.3.9）给出其概率，这就成了一个离散样本空间.

下面讨论的是更复杂一点的情况，即巴斯卡分布，它可以看作几何分布的一种推广.

4. 巴斯卡分布

考虑相继的伯努利试验，让我们考察要多长时间才会出现第

r 次成功.

若第 r 次成功发生在第 ζ 次试验，则必然有 $\zeta \geqslant r$.

我们以 C_k 表示第 r 次成功发生在第 k 次试验这一事件，并以 $f(k;r,p)$ 记其概率，C_k 发生当且仅当前面的 $k-1$ 次试验中有 $r-1$ 次成功，~~$k-r$ 次失败，~~而第 k 次试验的结果为成功，这些事件的概率分别为 $\binom{k-1}{r-1}p^{r-1}q^{k-r}$ 与 p，于是利用试验的独立性，得到

$$P(C_k)=\binom{k-1}{r-1}p^{r-1}q^{k-r}\cdot p=\binom{k-1}{r-1}p^rq^{k-r}$$

即

$$f(k;\ r,\ p)=\binom{k-1}{r-1}p^rq^{k-r},\quad k=r,\ r+1,\ \cdots \tag{2.3.11}$$

注意到

$$\begin{aligned}\sum_{k=r}^{\infty}f(k;\ r,\ p)&=\sum_{k=r}^{\infty}\binom{k-1}{r-1}p^rq^{k-r}\\&=\sum_{l=0}^{\infty}\binom{r+l-1}{r-1}p^rq^l=\sum_{l=0}^{\infty}\binom{-r}{l}(-1)^lp^rq^l\\&=p^r\ (1-q)^{-r}=1\end{aligned} \tag{2.3.12}$$

这里利用了推广的二项系数公式 $\binom{-r}{l}=(-1)^l\binom{r+l-1}{l}$.

$f(k;r,p)$ 称为**巴斯卡分布**. 特别当 $r=1$ 时，我们得到几何分布.

巴斯卡分布与著名的分赌注问题有关. 这个问题来源甚古，但直到德・梅尔向巴斯卡提出并导致巴斯卡与费马的通信，后又引起惠更斯（Huygens）的兴趣，才由他们三人分别给出正确答案. 巴斯卡和费马在解题中归结到取胜的概率，而惠更斯则引入数学期望的概念. 可以说这宣告了概率论这一学科的诞生.

分赌注问题大意如下：甲、乙两个赌徒按某种方式下注赌博，说定先胜 t 局者将赢得全部赌注，但进行到甲胜 r 局，乙胜 s 局 ($r<t$，$s<t$) 时，因故不得不中止，试问如何分配这些赌注才公平合理？

有人建议用已胜局数作比例分配赌注，即以 $r:s$ 来分配，但这种分法显然没有考虑到最终取胜的概率．若以 $n=t-r$ 及 $m=t-s$ 分别记甲及乙为达到最后胜利所须再胜的局数，又设甲在每局中取胜的概率为 p，我们便可以把分赌注问题归结为如下概率问题：在伯努利试验中，求在出现 m 次 $\bar{A}$ 之前出现 n 次 A 的概率．

若以 $p_甲$ 记上述概率，则它为甲最终取胜的概率，那么赌注以 $p_甲:1-p_甲$ 分配是公平合理的．巴斯卡和费马都在某种程度上达到这个结果．

现在，若利用巴斯卡分布，则容易写出答案

$$p_甲=\sum_{k=0}^{m-1}\binom{n+k-1}{k}p^nq^k \qquad (2.3.13)$$

或

$$p_甲=\sum_{k=n}^{\infty}\binom{m+k-1}{k}p^kq^m \qquad (2.3.14)$$

另外，容易证明，再赌 $n+m-1$ 局一定可以决定胜负．因此甲为取得最终胜利只须而且必须在后继的 $n+m-1$ 局中至少胜 n 局．这样，利用二项分布可以知道，

$$p_甲=\sum_{k=n}^{n+m-1}\binom{n+m-1}{k}p^kq^{n+m-1-k} \qquad (2.3.15)$$

可以证明上述三个答案是一致的（习题 30）．

下面是与巴斯卡分布有关的另一个有名例子．

［例 4］（巴拿赫火柴盒问题）数学家的左、右衣袋中各放有一盒装有 N 根火柴的火柴盒，每次抽烟时任取一盒用一根，求发

现一盒用光时，另一盒有 r 根的概率.

[解] 看作 $p=\frac{1}{2}$ 的伯努利试验. 要左边空而右边剩 r 根，应该是左边摸过 $N+1$ 次（前 N 次用去 N 根火柴，最后一次发觉火柴盒是空的），而右边摸过 $N-r$ 次，这事件的概率为

$$f\left(2N-r+1;\ N+1,\ \frac{1}{2}\right)=\binom{2N-r}{N}\left(\frac{1}{2}\right)^{2N-r+1}$$

对于右边先空的情况可同样考虑，因此所求的概率为

$$u_r=2\cdot f\left(2N-r+1;\ N+1,\ \frac{1}{2}\right)=\binom{2N-r}{N}2^{-2N+r}$$

三、直线上的随机游动

考虑 x 轴上的一个质点，假定它只能位于整数点，在时刻 $t=0$ 时，它处于初始位置 a（a 是整数），以后每隔单位时间，它总受到一个外力的随机作用，使位置发生变化，分别以概率 p 及概率 $q=1-p$ 向正的或负的方向移动一个单位，我们所关心的是质点在时刻 $t=n$ 时的位置. 用这种方式描述的质点运动称为**随机游动**.

若质点可以在整个数轴的整数点上游动，则称这种随机游动为**无限制随机游动**. 若在某点 d 设有一个吸收壁，质点一到达这点即被吸收而不再游动，因而整个游动也就结束了，这种随机游动称为**在 d 点有吸收壁的随机游动**. 此外还可以考虑带有反射壁及弹性壁的随机游动. 在一个随机游动中还可以具有不止一个壁.

当 $p=q=\frac{1}{2}$ 时，随机游动称为**对称的**，这时质点向左或向右移动的可能性相等.

自然科学中的大量问题归结为随机游动问题，例如随机游动模型可以作为布朗运动的初步近似. 股票价格涨落和汇率变化是否具有随机游动特征，更是现代金融界最充满火药味的论争之一. 概率论中的一些古典问题也引导到随机游动问题，事实上，随机

游动可以看作是伯努利试验的一种描述法.

关于随机游动，已进行过许多研究，我们只介绍它的两个最简单的模型.

无限制随机游动 假定质点在时刻0从原点出发，以 S_n 记它在时刻 $t=n$ 时的位置. 为了使质点在时刻 $t=n$ 时位于 k (k 也可以是负整数)，必须而且只须在前 n 次游动中向右游动的次数比向左游动的次数多 k 次，若以 x 记它在前 n 次游动中向右游动的次数，y 记向左移动的次数，则

$$\begin{cases} x+y=n \\ x-y=k \end{cases}$$

即 $x=\frac{n+k}{2}$，因为 x 是整数，所以 k 必须与 n 具有相同的奇偶性.

事件 $\{S_n=k\}$ 发生相当于要求在前 n 次游动中有 $\frac{n+k}{2}$ 次向右，$\frac{n-k}{2}$ 次向左，利用二项分布即得

$$P\{S_n=k\}=\binom{n}{\frac{n+k}{2}}q^{\frac{n-k}{2}}p^{\frac{n+k}{2}}$$

当 k 与 n 奇偶性相反时，概率为0.

两端带有吸收壁的随机游动 假定质点在时刻 $t=0$ 时，位于 $x=a$，而在 $x=0$ 及 $x=a+b$ 处各有一个吸收壁，我们来求质点在 $x=0$ 被吸收或在 $x=a+b$ 被吸收的概率. 用的是差分方程法.

若以 q_n 记质点的初始位置为 n 而最终在 $a+b$ 点被吸收的概率. 显然

$$q_0=0,\quad q_{a+b}=1 \tag{2.3.16}$$

如果某时刻质点位于 $x=n$，这里 $1\leqslant n\leqslant a+b-1$，则它要被 $x=a+b$ 吸收，有两种方式来实现：一种是接下去一次移动是向右的而最终被 $x=a+b$ 吸收；另一种是接下去一次移动是向左的而最终被 $x=a+b$ 吸收. 所以按全概率公式有

$$q_n = pq_{n+1} + qq_{n-1}, \quad n=1, 2, \cdots, a+b-1 \qquad (2.3.17)$$

这样，我们得到了关于 q_n 的一个二阶差分方程（2.3.17），再用边界条件（2.3.16）就可以求解. 利用这个差分方程系数的特殊性，比较方便的解法是把（2.3.17）改写成

$$p(q_{n+1}-q_n) = q(q_n - q_{n-1}), \quad n=1, 2, \cdots, a+b-1$$

若记 $c_n = q_{n+1} - q_n$，$r = \dfrac{q}{p}$，则又能写成

$$c_n = rc_{n-1}, \quad n=1, 2, \cdots, a+b-1$$

下面分两种情况求解：

(i)　$r=1$，即 $p=q=\dfrac{1}{2}$，也即对称随机游动的场合. 这时 $c_n = c_{n-1}$，因此，若记

$$q_{n+1} - q_n = q_n - q_{n-1} = \cdots = q_1 - q_0 = d$$

则

$$q_n = q_0 + nd$$

由于 $q_0 = 0$，$q_{a+b} = 1$，故有

$$q_n = \frac{n}{a+b}$$

特别地

$$q_a = \frac{a}{a+b} \qquad (2.3.18)$$

(ii)　$r \neq 1$，即 $p \neq q$ 的场合.

这时

$$c_n = rc_{n-1} = r^2 c_{n-2} = \cdots = r^n c_0$$

从而

$$q_n - q_0 = \sum_{k=0}^{n-1} (q_{k+1} - q_k) = \sum_{k=0}^{n-1} c_k = \sum_{k=0}^{n-1} r^k c_0 = \frac{1-r^n}{1-r} c_0$$

由于 $q_0 = 0$，$q_{a+b} = 1$，故有

$$\frac{1-r^{a+b}}{1-r} c_0 = 1$$

因此

$$q_n=\frac{1-r^n}{1-r^{a+b}}$$

特别地

$$q_a=\frac{1-r^a}{1-r^{a+b}}=\frac{1-\left(\frac{q}{p}\right)^a}{1-\left(\frac{q}{p}\right)^{a+b}} \tag{2.3.19}$$

若以 p_n 记质点自 n 出发而在 0 点被吸收的概率，同样可以列出差分方程

$$p_n=pp_{n+1}+qp_{n-1}, \quad n=1, 2, \cdots, a+b-1$$

及边界条件

$$p_0=1, \quad p_{a+b}=0$$

类似地可以求得，当 $p=q=\frac{1}{2}$ 时

$$p_a=\frac{b}{a+b} \tag{2.3.20}$$

而在 $p\neq q$ 的场合

$$p_a=\frac{1-\left(\frac{p}{q}\right)^b}{1-\left(\frac{p}{q}\right)^{a+b}}=\frac{\left(\frac{q}{p}\right)^a-\left(\frac{q}{p}\right)^{a+b}}{1-\left(\frac{q}{p}\right)^{a+b}} \tag{2.3.21}$$

不管在什么场合，都有

$$p_a+q_a=1$$

也就是说随机游动的质点最终一定要被两个端点之一所吸收.

注意，（2.3.18）及（2.3.20）也可以通过（2.3.19）及（2.3.21），令 $p\to q$ 用洛必达法则而得到.

顺便指出，两端有吸收壁的随机游动与概率论发展史上有名的赌徒输光问题有密切联系. 这个问题是这样叙述的：甲、乙进行赌博，其赌本分别为 a 及 b，若每局赌注为 1，而甲、乙在每局中赢的概率分别为 p 及 q. 试求甲（或乙）把赌本输光的概率.

四、推广的伯努利试验与多项分布

二项分布可以容易地推广到 n 次重复独立试验且每次试验可能有若干个结果的情形. 把每次试验的可能结果记为 A_1, A_2, …, A_r, 而 $P(A_i)=p_i$, $i=1$, 2, …, r, 且

$$p_1+p_2+\cdots+p_r=1, \qquad p_i\geqslant 0 \tag{2.3.22}$$

当 $r=2$ 时, 我们得到伯努利试验.

在这种推广的伯努利试验中, 不难导出, 在 n 次试验中 A_1 出现 k_1 次, A_2 出现 k_2 次, ……A_r 出现 k_r 次的概率为

$$\frac{n!}{k_1!\ k_2!\ \cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r} \tag{2.3.23}$$

这里 $k_i\geqslant 0$, 且 $k_1+k_2+\cdots+k_r=n$.

公式 (2.3.23) 称为**多项分布**, 因为它是 $(p_1+p_2+\cdots+p_r)^n$ 的展开式的一般项, 而且由 (2.3.22) 知

$$\sum_{\substack{k_i\geqslant 0\\ k_1+k_2+\cdots+k_r=n}}\frac{n!}{k_1!\ k_2!\ \cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}=1 \tag{2.3.24}$$

显然多项分布是二项分布的推广. 二项分布中的许多结果都能平行地推广到多项分布的场合, 以后我们只详细讨论二项分布的有关问题.

在产品检查中,若对产品质量所用的标准不只是好品与废品, 而是分得更细, 例如有一等品, 二等品, 三等品, 等外品四类, 则从中取出 n 件,求一等品有 k_1 件,二等品有 k_2 件,三等品有 k_3 件, 等外品有 k_4 件的概率时便得到多项分布.

[例 5]　人类的血型分为 O, A, B, AB 四型, 假定某地区的居民中这四种血型的人的百分比分别为 0.4, 0.3, 0.25, 0.05, 若从此地区居民中随机地选出 5 人, 求有两个为 O 型, 其他三个分别是 A, B, AB 型的概率.

[解]　推广的伯努利试验可以用于这个场合, 所求的概率为

$$P=\frac{5!}{2!\ 1!\ 1!\ 1!}\times 0.4^2\times 0.3\times 0.25\times 0.05=0.036$$

也可以研究平面上或空间的随机游动，下面是简单的一例.

［例 6］（平面上随机游动）一质点从平面上某点出发，等可能地向上，下，左，右方向移动，每次移动的距离为 1，求经过 $2n$ 次移动后回到出发点的概率.

［解］ 这可以归结为上述推广的伯努利试验的问题，分别以事件 A_1，A_2，A_3，A_4 表示质点向上，下，左，右移动一格，则 $p_1=p_2=p_3=p_4=\frac{1}{4}$. 若要在 $2n$ 次移动后回到原来的出发点，则向左移动的次数与向右移动的次数应该相等，向上移动的次数与向下移动的次数也应该相等. 而总移动次数为 $2n$，因此所求概率为

$$P=\sum_{k+m=n}\frac{(2n)!}{(k!)^2(m!)^2}\left(\frac{1}{4}\right)^{2n}=\sum_{k=0}^{n}\frac{(2n)!}{(k!)^2[(n-k)!]^2}\left(\frac{1}{4}\right)^{2n}$$

$$=\left(\frac{1}{4}\right)^{2n}\frac{(2n)!}{(n!)^2}\sum_{k=0}^{n}\left[\frac{n!}{k!\ (n-k)!}\right]^2$$

$$=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}\sum_{k=0}^{n}\binom{n}{k}^2=\left(\frac{1}{4}\right)^{2n}\binom{2n}{n}^2$$

最后一个等式用到（1.3.7）.

§4. 二项分布与泊松分布

一、二项分布的性质及计算

1. 二项分布的计算

在上一节中我们导出了在 n 次伯努利试验中正好出现 k 次成功的概率 $b(k;n,p)$：

$$b(k;n,p)=\binom{n}{k}p^kq^{n-k},\quad k=0,1,2,\cdots,n \qquad (2.4.1)$$

其中 $q=1-p$.

$b(k;n,p)$，$k=0,1,2,\cdots,n$ 称为**二项分布**，在概率论中占有很重要的地位．由于在许多实际问题中出现二项分布，并且要计算其数值，因此讨论二项分布的计算显得非常重要．

二项分布有现成的表可查，这种表对不同的 n 及 p 给出了 $b(k;n,p)$ 的数值．

为了增加对二项分布的感性认识及计算的需要，我们选取了下列二项分布数值表．

表 2.4.1 中给出了对于 $n=20$ 及 $p_1=0.1$，$p_2=0.3$，$p_3=0.5$ 的二项分布数值表．

二项分布表只对 $p\leqslant 0.5$ 给出，因为对于 $p>0.5$ 的概率不难经下式计算得到：

$$b(k;n,p)=b(n-k;n,1-p) \tag{2.4.2}$$

表 2.4.1 二项分布数值表

k	$b(k;20,p)$			k	$b(k;20,p)$		
	$p_1=0.1$	$p_2=0.3$	$p_3=0.5$		$p_1=0.1$	$p_2=0.3$	$p_3=0.5$
0	0.1216	0.0008	—	11	—	0.0120	0.1602
1	0.2702	0.0068	—	12	—	0.0039	0.1201
2	0.2852	0.0278	0.0002	13	—	0.0010	0.0739
3	0.1901	0.0716	0.0011	14	—	0.0002	0.0370
4	0.0898	0.1304	0.0046	15	—	—	0.0148
5	0.0319	0.1789	0.0148	16	—	—	0.0046
6	0.0089	0.1916	0.0370	17	—	—	0.0011
7	0.0020	0.1643	0.0739	18	—	—	0.0002
8	0.0004	0.1144	0.1201	19	—	—	—
9	0.0001	0.0654	0.1602	20	—	—	—
10	—	0.0308	0.1762				

当 $p>0.5$ 时，$1-p<0.5$，仍能利用分布数值表．

为了对二项分布的变化情况有个直观了解，我们把表 2.4.1 中的几个分布用图 11 表示出来．

从图中可以看出，对于固定的 n 及 p，当 k 增加时，$b(k;n,p)$ 先随之增加并达到某极大值，以后又下降．此外，当概率 p 越

与$\frac{1}{2}$接近时，分布越接近对称．

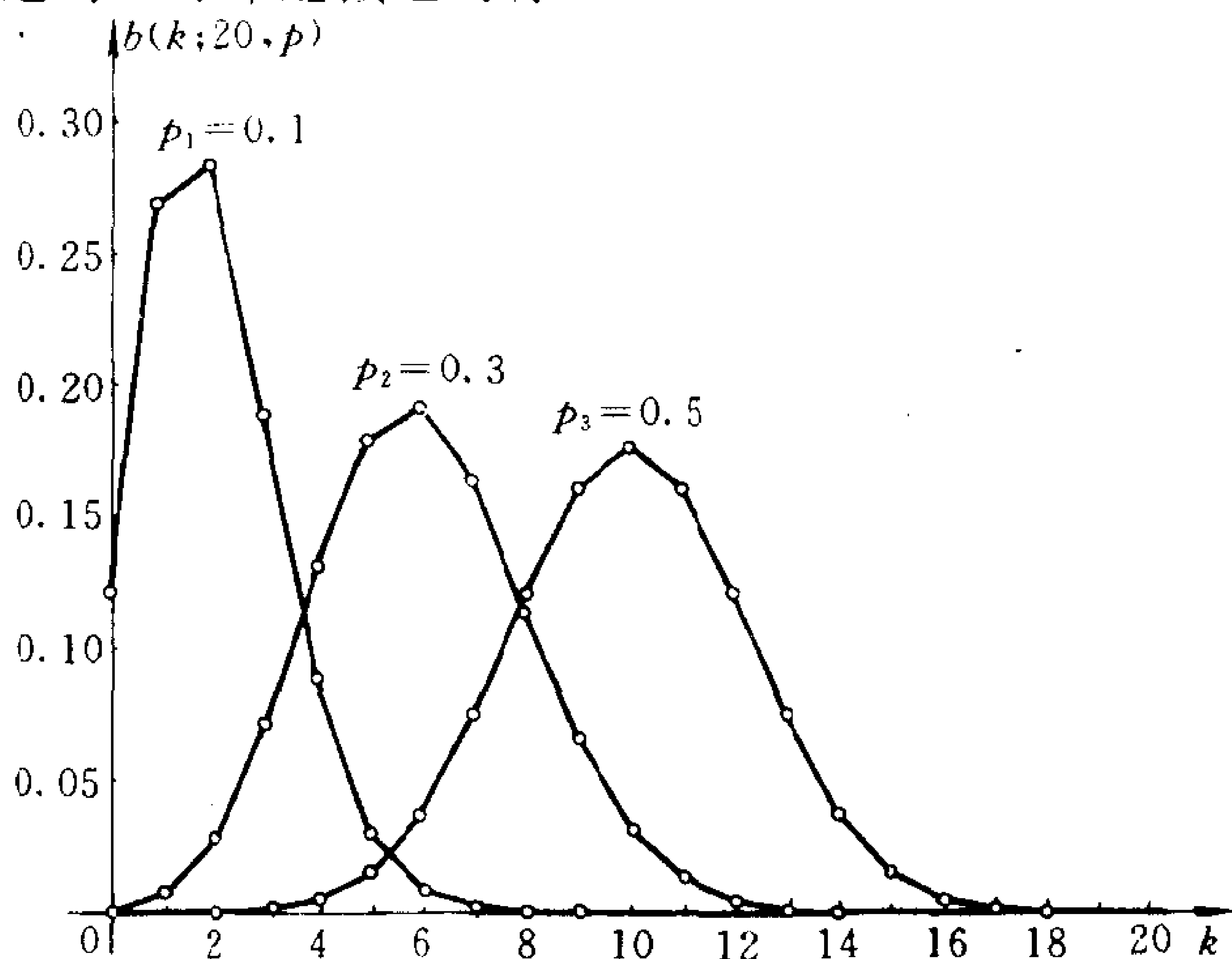


图 11　二项分布图

［例 1］　一大批电子管中有 10% 已损坏，若我们从这批电子管中随机地选取 20 只来组成一个线路，问这线路能正常工作（即所选的 20 个电子管全部是好的）的概率是多少？

［解］　因为这批电子管数量很大，因此可近似地把选取 20 个管作为进行独立试验．这是一个伯努利概型问题，若把选得好电子管作为成功．其概率为 0.9，则所求的概率为

$$b(20;20,0.9)=\binom{20}{20}0.9^{20}=0.9^{20}$$

这个数值可以利用对数表计算，但更方便的是利用表 2.4.1 计算，由（2.4.2），

$$b(20;20,0.9)=b(0;20,0.1)$$

利用表 2.4.1 知所求的概率为 0.1216.

［例 2］　（血清的试验）设在家畜中感染某种疾病的概率是 30%，新发现了一种血清可能对预防此病有效，为此对 20 只健康的动物注射这种血清．若注射后只有一只动物受感染，我们应对此种血清的作用作何评价？

假如血清毫无价值，那么注射后的动物受感染的概率还是30%，则这 20 只动物中有 k 只受感染的概率为b（k；20，0.3）.

发生只有一只动物受感染或更好的情况（无动物受感染）的概率为

$$b(0;20,0.3)+b(1;20,0.3)=0.0008+0.0068=0.0076$$

这个概率如此之小，因此我们不能认为血清毫无价值.

如果注射后的 20 只动物中有 4 只受感染，我们是否相信此种血清有效？这个问题留给读者思考.

这里的做法是：先依照我们关心的问题提出一个假设，然后用实验得出的数据，利用概率论方法，计算某个事件在假设成立下的概率，最后根据这概率的大小来决定是接受还是拒绝原来的假设，这是数理统计中有名的**统计假设检验法**.

2. 二项分布的性质

我们来考察 b（k；n，p）随 k 及 n 变化的情况. 从图 11 可以看出，当 n 固定时，b（k；n，p）先随 k 增加而增大，达到某一极大值后又逐渐下降. 现在对它进行严格讨论.

由于对 $0<p<1$,

$$\frac{b(k;n,p)}{b(k-1;n,p)}=\frac{(n-k+1)p}{kq}=1+\frac{(n+1)p-k}{kq}$$

因此

当 $k<(n+1)p$ 时，$b(k;n,p)>b(k-1;n,p)$

当 $k=(n+1)p$ 时，$b(k;n,p)=b(k-1;n,p)$

当 $k>(n+1)p$ 时，$b(k;n,p)<b(k-1;n,p)$

因为$(n+1)p$ 不一定是整数，而二项分布中的 k 只取整数值，所以存在整数 m，使得（$n+1$）$p-1<m\leqslant$（$n+1$）p，而且当 k 从 0 变到 n 时，b（k；n，p）起先单调上升，当 $k=m$ 时达到极大值，后来又单调下降. 但若（$n+1$）$p=m$ 则这时 b（m；n，p）= b（$m-1$；n，p）同时达到极大值.

使 $b(k$；n，$p)$ 取最大值的项 $b(m$；n，$p)$ 称为 $b(k$；n，$p)$

的**中心项**，而 m 称为**最可能成功次数**．由上面讨论知 $m=[(n+1)p]$（即 m 是 $(n+1)p$ 的整数部分）．若 $(n+1)p$ 是整数，则 $m-1$ 亦为最可能成功次数．

［例 3］ 设每颗子弹打中飞机的概率为 0.01，问在 500 发子弹中打中飞机的最可能次数是多少？并求其相应的概率．

［解］ 这是伯努利概型，打中飞机子弹数服从二项分布．$n=500$，$p=0.01$，$(n+1)p=5.01$，$[(n+1)p]=5$．所以最可能成功次数为 5．相应概率为

$$b(5;500,0.01)=\binom{500}{5}(0.01)^5\cdot(0.99)^{495}=0.17635 \tag{2.4.3}$$

应该注意，若 $0<p<1$，则当 n 值相当大时，即便是最可能成功次数 m 发生的概率也相当小，对于其他的 k，则 $b(k;n,p)$ 自然更小了，以后将会看到最可能成功次数 m 发生的概率接近于

$$(2\pi npq)^{-\frac{1}{2}}$$

（当 n 相当大时），因此当 $n\to\infty$ 时，这概率趋于 0.

3．产品抽样验收与 (n,c) 方案

由于生产过程总有种种无法完全控制的因素，因此工艺规范也允许加工的尺寸有一定的公差，或允许产品中含有少量废品，这事实上是承认生产过程的随机性．

在产品质量管理中，全面检验一般是不可能的，因此采用抽样检查的办法．

抽样检验若用于生产过程中，则成为在线生产过程质量管理的一部分，此外就是用于产品的验收．

如果每个产品要么是好品要么是废品，这时关心的是废品数或废品率，这是计数抽样验收中最简单的情况．

对质量的要求大体上可以归结为：存在 p_0 及 p_1 满足 $0<p_0\leqslant p_1<1$，当废品率 $p\leqslant p_0$ 时，接收这批产品；而当 $p\geqslant p_1$ 时，拒绝这批产品．

最简单也是最基本的验收方案是：抽 n 件产品进行检验，当废品数 $\leqslant c$ 时，接收该批产品；否则拒绝.

这个方案称为（n，c）**方案**.

由于抽样的随机性，任何验收方案都可能犯两类错误：其一，拒收一批合格品；其二，接收一批不合格品．前者为生产者风险；后者为消费者风险．当然希望减小这两类风险，即降低犯两类错误的概率．这也为比较两种不同验收方案的优劣提供了客观的标准.

为刻画验收方案的性能，一般引进 $L(p)$，它表示当废品率为 p 时，接收该批产品的概率．若以 p 为横坐标，$L(p)$ 为纵坐标作图，则所得的曲线称为**操作特性曲线**(operating characteristic curve)，简称 **OC 曲线**（见图 12）.

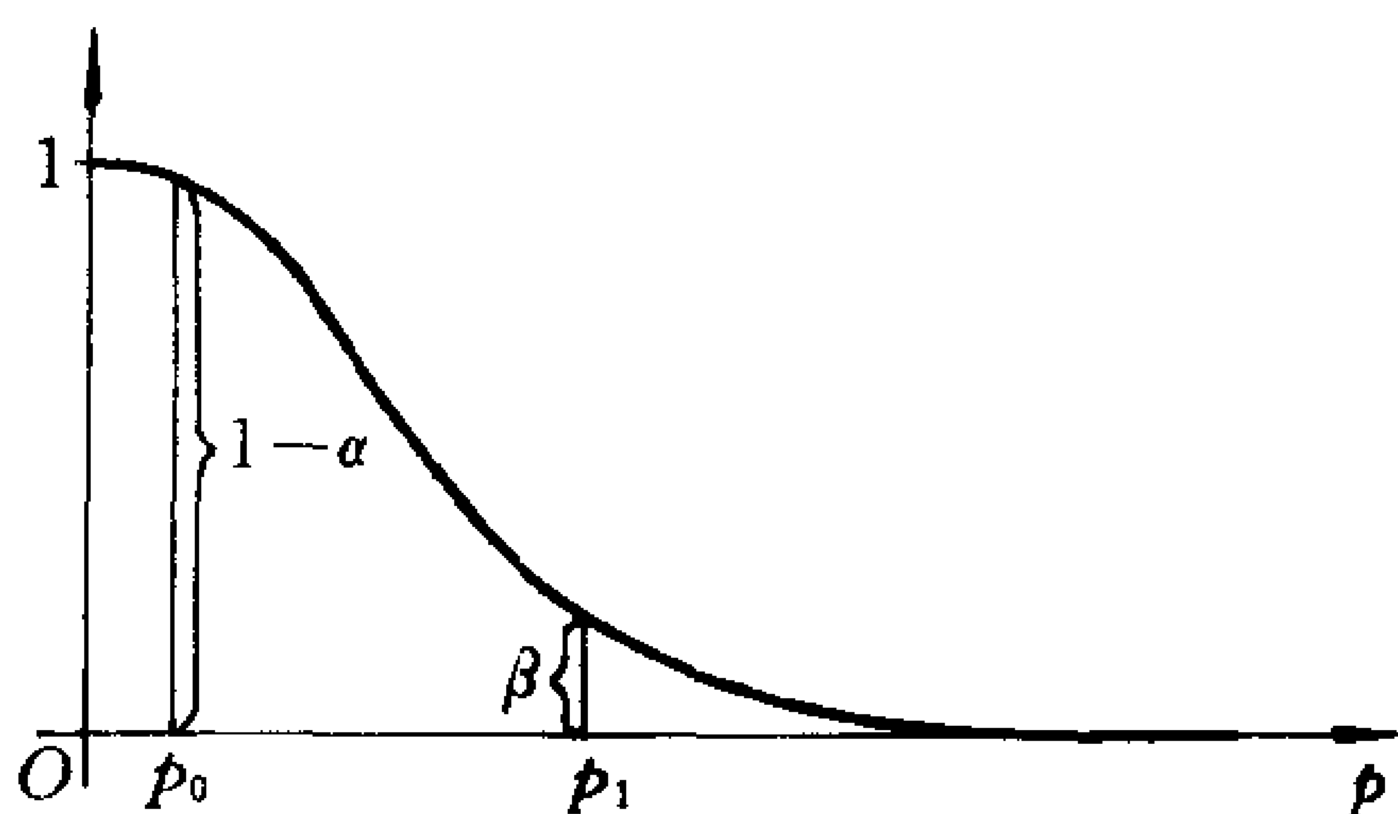


图 12 OC 曲线

对（n，c）方案而言，若抽样是放回的，则利用二项分布容易得到

$$L(p)=\sum_{k=0}^{c}\binom{n}{k}p^{k}(1-p)^{n-k} \tag{2.4.4}$$

因此，问题归结为找 n 及 c，使得

$$\begin{cases} L(p)\geqslant 1-\alpha, & \text{当 } p\leqslant p_0 \text{ 时} \\ L(p)\leqslant \beta, & \text{当 } p>p_1 \text{ 时} \end{cases} \tag{2.4.5}$$

这里 α，β 是两个不大的正数，按需要给定.

理想的验收方案要求 $\alpha=\beta=0$，这是无法实现的，但可作为比较的基准.

4. 应用的例子

我们举一些应用的例子，说明二项分布的重要性，同时也提出一些问题.

[例 4]　保险事业是最早使用概率论的部门之一. 保险公司为了决定保险金数额，估算公司的利润和破产的风险，需要计算各种各样的概率. 下面是典型问题之一. 根据生命表知道，某年龄段保险者里，一年中每个人死亡的概率等于 0.005，现有 10000 个这类人参加人寿保险，试求在未来一年中在这些保险者里面，(1) 有 40 个人死亡的概率；(2) 死亡人数不超过 70 个的概率.

[解]　作为初步近似，可以利用伯努利概型，$n=10000$，$p=0.005$，设μ为未来一年中这些人里面死亡的人数，则所求的概率分别为

$$(1)\ b(40;10000,0.005)=\binom{10000}{40}(0.005)^{40}(0.995)^{9960} \tag{2.4.6}$$

$$(2)\ P\{\mu\leqslant 70\}=\sum_{k=0}^{70}b(k;10000,0.005)=\sum_{k=0}^{70}\binom{10000}{k}(0.005)^{k}(0.995)^{10000-k} \tag{2.4.7}$$

直接计算这些数值相当困难，要有更好的计算方法.

[例 5]　甲、乙两容器，容量各为 1 升，每个各含 2.7×10^{22} 个气体分子，现将两容器接触，经过相当长的时间后$\left(\text{即这时每个分子落在两容器中的概率各为}\frac{1}{2}\right)$，求两容器中分子数之差超过分子总数的 100 亿分之一的概率.

[解]　分子总数为 5.4×10^{22}，其一百亿分之一（即 10^{-10}）为 5.4×10^{12}，设混和后甲容器的分子数为 μ，则乙容器的分子数为

$5.4\times10^{22}-\mu$，现要求事件

$$|\mu-(5.4\times10^{22}-\mu)|>5.4\times10^{12}$$

即事件

$$|\mu-2.7\times10^{22}|>2.7\times10^{12} \tag{2.4.8}$$

的概率，由二项分布

$$P\{|\mu-2.7\times10^{22}|>2.7\times10^{12}\}$$
$$=\sum_k b\left(k;5.4\times10^{22},\frac{1}{2}\right) \tag{2.4.9}$$

上式和号对一切满足$|k-2.7\times10^{22}|>2.7\times10^{12}$的 k 求和，这概率是无法直接计算的．

从上面两个例子中可以看到，计算二项分布的数值时，由于试验次数 n 经常很大，因此实际计算 $b(k;n,p)$ 及 $\sum_k b(k;n,p)$ 都很困难，有时甚至不可能．例如（2.4.9）中须算的项数有 5.4×10^{12}之多，逐项计算是不可能的．

在这种情况下，寻找更有效的计算法是必要的，即便是近似公式也好．这可以利用概率论中的极限定理来实现，关于极限定理的讨论将在第五章进行．

二、二项分布的泊松逼近

在很多应用问题中，我们常常遇到这样的伯努利试验，其中，相对地说，n 大，p 小，而乘积 $\lambda=np$ 大小适中．在这种情况下，有一个便于使用的近似公式，下面我们来推导它．

定理 2.4.1（泊松） 在独立试验中，以 p_n 代表事件 A 在试验中出现的概率，它与试验总数 n 有关，如果 $np_n\to\lambda$，则当 $n\to\infty$ 时，

$$b(k;n,p_n)\to\frac{\lambda^k}{k!}e^{-\lambda}$$

［证明］ 记 $\lambda_n=np_n$，则

$$b(k;n,p_n)=\binom{n}{k}p_n^k(1-p_n)^{n-k}$$

$$=\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda_n}{n}\right)^k\left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

$$=\frac{\lambda_n^k}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)\left(1-\frac{\lambda_n}{n}\right)^{n-k}$$

由于对固定的 k 有

$$\lim_{n\to\infty}\lambda_n^k=\lambda^k,\lim_{n\to\infty}\left(1-\frac{\lambda_n}{n}\right)^{n-k}=e^{-\lambda}$$

及

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)=1$$

因此

$$\lim_{n\to\infty}b(k;n,p_n)=\frac{\lambda^k}{k!}e^{-\lambda}$$

定理证毕.

$$p(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda},k=0,1,2,\cdots \qquad (2.4.10)$$

称为**泊松分布**（**Poisson distribution**），λ 称为它的参数.

特别地

$$\sum_{k=0}^{\infty}p(k;\lambda)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}\cdot e^{\lambda}=1 \qquad (2.4.11)$$

泊松分布是概率论中很重要的一种分布.

在应用中，当 p 相当小（一般当 $p\leqslant 0.1$）时，我们用下面近似公式

$$b(k;n,p)\approx\frac{(np)^k}{k!}e^{-np} \qquad (2.4.12)$$

［例 6］ 假如生三胞胎的概率为 10^{-4}，求在 100000 次生育中，有 0，1，2 次生三胞胎的概率.

［解］ 这可以看作伯努利试验；$n=100000$，$p=0.0001$，所求的概率直接计算为

$$b(0;100000,0.0001)=0.000045378$$
$$b(1;100000,0.0001)=0.00045382$$
$$b(2;100000,0.0001)=0.0022693$$

这时也可用泊松逼近，$\lambda=np=10$，而

$$p(0;10)=0.00004540$$
$$p(1;10)=0.0004540$$
$$p(2;10)=0.002270$$

可见近似程度很令人满意.

图13给出了泊松分布逼近二项分布的一个图示，吻合程度甚好.

三、泊松分布

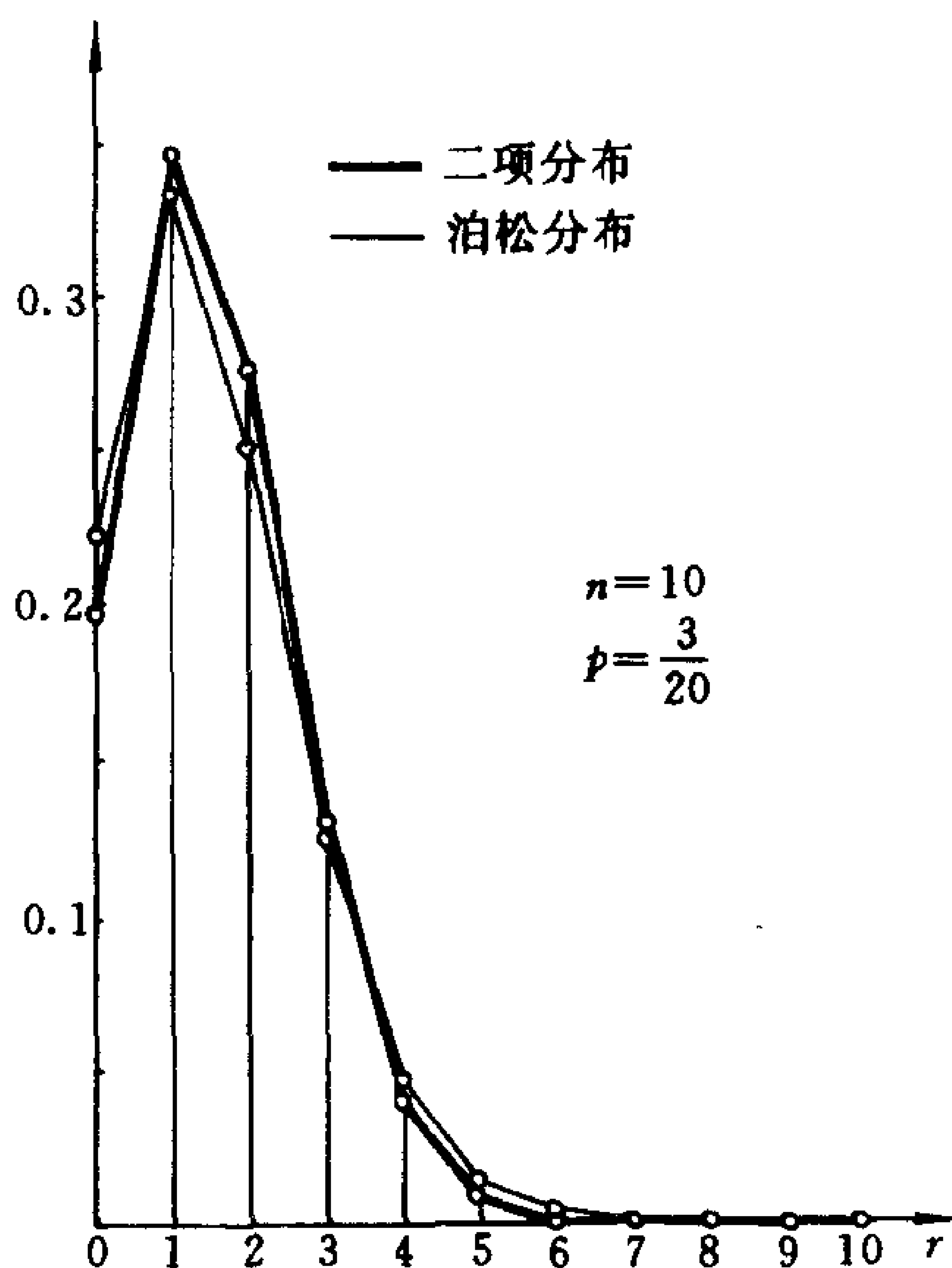


图13 二项分布与泊松分布

在历史上泊松分布是作为二项分布的近似，于1837年由法国数学家泊松引入的. 近数10年来，泊松分布日益显示其重要性，成了概率论中最重要的几个分布之一. 其原因主要是下面两点.

首先是已经发现许多随机现象服从泊松分布. 这种情况特别集中在两个领域中. 一是社会生活，对服务的各种要求，诸如电话交换台中来到的呼叫数，公共汽车站来到的乘客数等等都近似地服从泊松分布，因此在运筹学及管理科学中泊松分布占有很突出的地位；另一领域是物理科学，放射性分裂落到某区域的质点数，热电子的发射，显微镜下落在某区域中的血球或微生物的数目等等都服从泊松分布.

其次，对泊松分布的深入研究（特别是通过随机过程的研究）已发现它具有许多特殊的性质和作用，打个不很恰当的譬喻，似乎泊松分布是构造随机现象的“基本粒子”之一.

图 14 是对于不同 λ 值的泊松分布图. 为了计算泊松分布的数值，有许多专门的表格可供查用. 本书附录二中也附有这样的表.

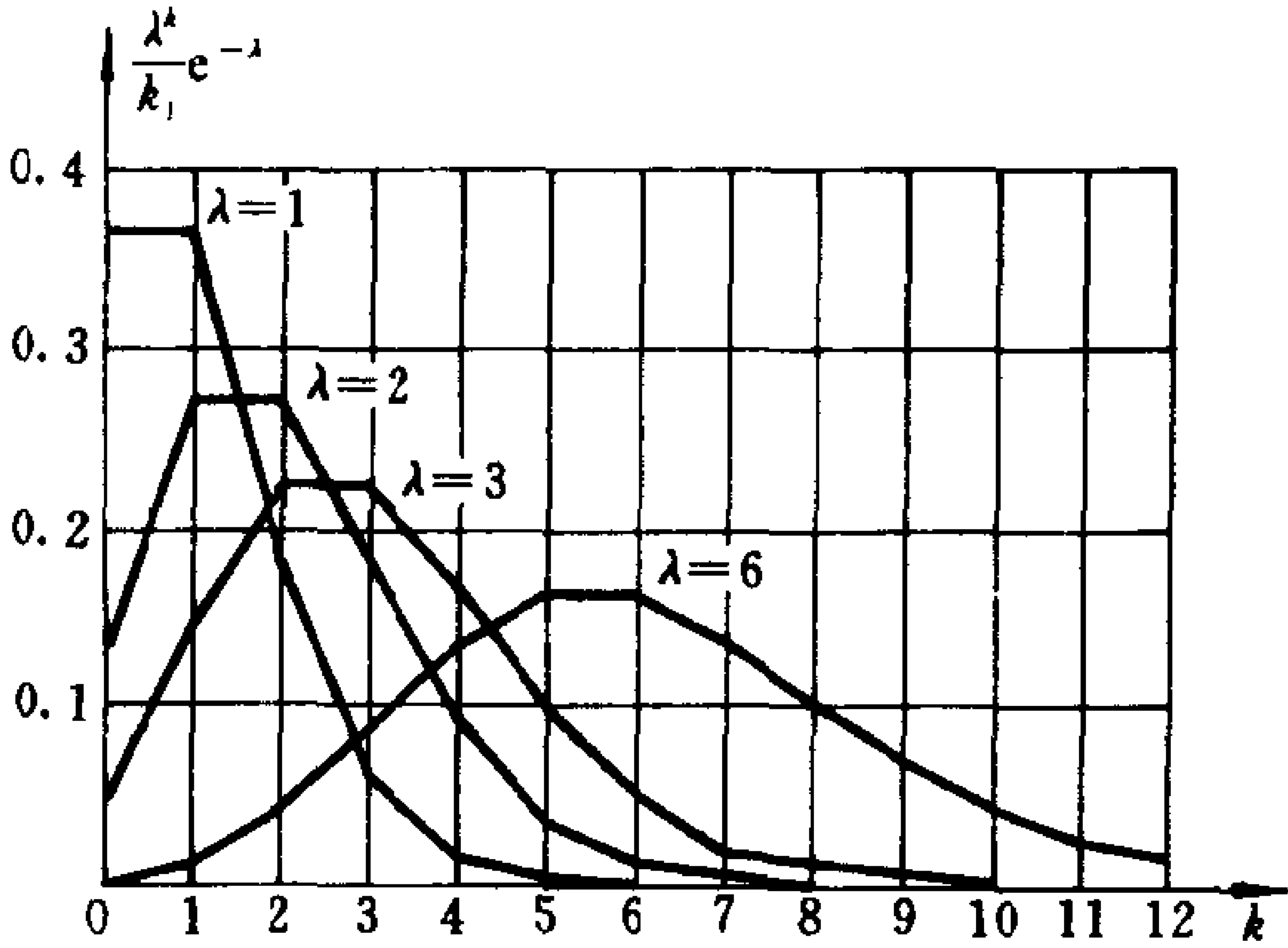


图 14 泊松分布图

下面提供两个有关的统计资料作为例子.

[例 7] 对上海市某公共汽车站的客流进行调查，统计了某天上午 10：30 至 11：47 左右每隔 20 秒钟来到的乘客批数（每批可能有数人同时来到），共得 230 个记录，分别计算了来到 0 批，1 批，2 批，3 批，4 批及 4 批以上乘客的时间区间的频数，结果列于表 2.4.2 中，其相应的频率与 $\lambda=0.87$ 的泊松分布符合得很好.

表 2.4.2 公共汽车客流统计

来到批数 i	0	1	2	3	$\geqslant 4$	总 共
频数 n_i	100	81	34	9	6	230
频率 $f_i=\frac{n_i}{n}$	0.43	0.35	0.15	0.04	0.03	
$p_i=\frac{\lambda^i}{i!}e^{-\lambda}$	0.42	0.36	0.16	0.05	0.01	

［例 8］ 放射性物质放射出的 α 质点数是服从泊松分布的有名例子. 1910 年 Rutherford 等人的著名实验揭露了这个事实.

在这个实验中，观察了长为 7.5 秒的时间间隔里到达某指定区域的质点数，共观察了 $N=2608$ 次，表 2.4.3 给出观察值与理论值的对照，N_k 表示在 N 次观察中发生"在 7.5 秒内落到指定区域的质点数为 k"的观察次数，理论值是 N_p (k; 3.870)，理论值与实验值很近似.

表 2.4.3 Rutherford 实验理论值与实验值对照表①

k	N_k	N_p (k; 3.870)
0	57	54.399
1	203	210.523
2	383	407.361
3	525	525.496
4	532	508.418
5	408	393.515
6	273	253.817
7	139	140.325
8	45	67.882
9	27	29.189
$k\geqslant10$	16	17.075
总　计	2608	2608.000

在说明了泊松分布的常见性之后，我们转入介绍产生泊松分布的机制. 经过研究，已经弄清了服从泊松分布的条件. 为了便于理解，我们将结合电话呼叫流来叙述这个重要结果.

我们先证明一个以后屡次要用到的数学分析结论.

① 引自费勒. 概率论及其应用，离散空间（上册）. 科学出版社. 1964，第 155 页。

引理 2.4.1 若 $f(x)$ 是连续函数（或单调函数），且对一切 x，y（或一切 $x\geqslant 0$，$y\geqslant 0$）成立

$$f(x)f(y)=f(x+y) \tag{2.4.13}$$

则

$$f(x)=a^x \tag{2.4.14}$$

其中 $a\geqslant 0$，是某一常数.

［证明］由（2.4.13）知对任意 x，

$$f(x)=\left[f\left(\frac{x}{2}\right)\right]^2\geqslant 0$$

因此 $f(x)$ 非负. 反复使用（2.4.13），对任意正整数 n 及实数 x 有

$$f(nx)=[f(x)]^n$$

在上式中取 $x=\frac{1}{n}$，得

$$f(1)=\left[f\left(\frac{1}{n}\right)\right]^n$$

记 $a=f(1)\geqslant 0$，则

$$f\left(\frac{1}{n}\right)=a^{\frac{1}{n}}$$

因此，对任意正整数 m 及 n，成立

$$f\left(\frac{m}{n}\right)=\left[f\left(\frac{1}{n}\right)\right]^m=a^{\frac{m}{n}}$$

这样，我们已证得（2.4.14）对一切有理数成立，再利用连续性或单调性可以证明对无理数也成立，从而证明了引理.

［泊松过程］ 考虑来到某交换装置的电话呼叫数，假定它具有下面三个性质：

(i) **平稳性** 在 $[t_0, t_0+t)$ 中来到的呼叫数只与时间间隔长度 t 有关而与时间起点 t_0 无关. 若以 $P_k(t)$ 记在长度为 t 的时间区间中来到 k 个呼叫的概率，当然

$$\sum_{k=0}^{\infty}P_k(t)=1 \tag{2.4.15}$$

对任何 $t>0$ 成立.

过程的平稳性表示了它的概率规律不随时间的推移而改变.

(ii) **独立增量性（无后效性）** 在 $[t_0, t_0+t)$ 内来到 k 个呼叫这一事件与时刻 t_0 以前发生的事件独立. 换言之，在对时刻 t_0 以前的事件发生情况所作的任何假定之下，计算出来的在 $[t_0, t_0+t)$ 内发生 k 个呼叫的条件概率都等于同一事件的无条件概率. 独立增量性表明在互不相交的时间区间内过程进行的相互独立性.

(iii) **普通性** 在充分小的时间间隔中，最多来到一个呼叫. 即，若记

$$\psi(t)=\sum_{k=2}^{\infty}P_k(t)=1-P_0(t)-P_1(t) \tag{2.4.16}$$

应有 $\psi(t)=o(t)$，即

$$\lim_{t\to 0}\frac{\psi(t)}{t}=0 \tag{2.4.17}$$

普通性表明，在同一时间瞬间来两个或两个以上呼叫实际上是不可能的.

下面我们求 $P_k(t)$.

对 $\Delta t>0$，考虑 $[0, t+\Delta t)$ 中来到 k 个呼叫的概率 $P_k(t+\Delta t)$，由独立增量性及全概率公式

$$P_k(t+\Delta t)=P_k(t)P_0(\Delta t)+P_{k-1}(t)P_1(\Delta t)+\cdots$$
$$+P_0(t)P_k(\Delta t),k\geqslant 0 \tag{2.4.18}$$

（对 $n\geqslant 1$，假定 $P_{-n}(t)=0$.）

特别地

$$P_0(t+\Delta t)=P_0(t)P_0(\Delta t) \tag{2.4.19}$$

$P_0(t)$ 表示在长度为 t 的时间间隔中没有来呼叫的概率，因此它关于 t 单调下降，由引理 2.4.1 知

$$P_0(t)=a^t$$

其中 $a\geqslant 0$，若 $a=0$，则 $P_0(t)\equiv 0$，这说明在不管怎么短的时间

间隔内都要来呼叫，因此在有限时间间隔中要来无穷多个呼叫，这种情形不在我们的考虑之列．此外，因 P_0（t）是概率，故应有 $a\leqslant 1$，而当 $a=1$ 时，P_0（t）$\equiv 1$，这表明永不来呼叫，也不是我们感兴趣的情形，所以应有 $0<a<1$，从而存在 $\lambda>0$，使

$$P_0(t)=\mathrm{e}^{-\lambda t} \tag{2.4.20}$$

因此当 $\Delta t\to 0$ 时，我们有

$$P_0(\Delta t)=\mathrm{e}^{-\lambda\Delta t}=1-\lambda\Delta t+o(\Delta t)$$

$$P_1(\Delta t)=1-P_0(\Delta t)-\psi(\Delta t)=\lambda\Delta t+o(\Delta t)$$

$$\sum_{l=2}^{\infty}P_{k-l}(t)P_l(\Delta t)\leqslant\sum_{l=2}^{\infty}P_l(\Delta t)=\psi(\Delta t)=o(\Delta t)$$

故由（2.4.18）得

$$P_k(t+\Delta t)=P_k(t)(1-\lambda\Delta t)+P_{k-1}(t)\cdot\lambda\Delta t+o(\Delta t)$$

$$k\geqslant 1$$

因此

$$\frac{P_k(t+\Delta t)-P_k(t)}{\Delta t}=\lambda[P_{k-1}(t)-P_k(t)]+o(1)$$

$$k\geqslant 1$$

令 $\Delta t\to 0$，得

$$P_k'(t)=\lambda[P_{k-1}(t)-P_k(t)],\ k\geqslant 1$$

由于已知 $P_0(t)=\mathrm{e}^{-\lambda t}$，故有 $P_1'(t)=\lambda[\mathrm{e}^{-\lambda t}-P_1(t)]$，可解得 $P_1(t)=\lambda t\mathrm{e}^{-\lambda t}$，这样下去，可解得一切 $P_k(t)$．

$$P_k(t)=\frac{(\lambda t)^k}{k!}\mathrm{e}^{-\lambda t},\ k=0,\ 1,\ 2,\ \cdots$$

这正是泊松分布，参数为 λt．

在随机过程理论中，这里所得的结果及所用的方法将被大大推广．

第二章小结

本章中我们讨论了（统计）独立性的概念，它是概率论中最

重要的概念之一．独立性是概率论特有的概念，它的引进大大推动了概率论的发展，前期概率论中最重要的一些结果大都是在独立性的假定下获得的，只有到了近代才开始研究一些不独立但常在另一种较弱独立性假定下的概率模型．读者能在本书的后面部分充分体验到独立性的重要．

我们定义了事件的独立性与试验的独立性，读者应把它们进行对比．此外，两个事件独立性的定义与多个事件独立性的定义形式上也有区别，应理解为什么会发生这种情况．

条件概率是概率论中另一重要概念，它与独立性有密切联系．在不具有独立性的场合，它将扮演重要角色．条件概率也是某种概率．

本章还导出几个基本公式：乘法公式及其推广形式，全概率公式，贝叶斯公式，它们以后经常被用到．

伯努利试验是概率论中最重要的概型之一．正是通过对这个概型的不断深入地研究，逐渐提出了概率论特有的课题，创造出相应的工具与方法．它对后来的发展有着不可估量的影响．伯努利试验概型在应用上也很重要．在第五章，我们将继续讨论这个概型，证明有关的极限定理并最后解决一些本章遗留的计算问题．

二项分布与泊松分布是概率论中最重要的三个分布中的两个，另一个分布——正态分布，将在下章出现．这两个分布在理论研究和实际应用中都很重要．二项分布广泛应用于抽样检查，而泊松分布则大量出现于社会生活和物理现象中．泊松过程的结构分析是随机过程理论的最基本成果之一．

到目前为止，本课程的展开基本上还是与概率论的历史发展相平行的，而且，我们已经为概率论另一重要概念——随机变量的引入作了不少准备．

习　题　二

1．把字母 M、A、X、A、M 分别写在一张卡片上，充分混和后重新排

列，问正好得到顺序 MAXAM 的概率是多少？

2. 有三个孩子的家庭中，已知有一个是女孩，求至少有一个男孩的概率.

3. 若 M 件产品中包含 m 件废品，今在其中任取两件，求：(1) 取出的两件中至少有一件是废品的概率；(2) 已知取出的两件中有一件是废品的条件下，另一件也是废品的条件概率；(3) 已知两件中有一件不是废品的条件下，另一件是废品的条件概率.

4. 证明：对于事件 A，B，关系式

$$P^2(AB)+P^2(\bar{A}B)+P^2(A\bar{B})+P^2(\bar{A}\ \bar{B})=\frac{1}{4}$$

成立的充要条件为

$$P(A)=P(B)=\frac{1}{2},P(AB)=\frac{1}{4}$$

5. 袋中有 a 只黑球，b 只白球，甲、乙、丙三人依次从袋中取出一球（取后不放回），试分别求出三个人各自取得白球的概率.

6. 甲袋中有 a 只白球，b 只黑球，乙袋中有 α 只白球，β 只黑球，某人从甲袋中任取两球投入乙袋，然后在乙袋中任取两球，问最后取出的两球全为白球的概率是多少？

7. 设有 N 个袋子，每个袋子中装有 a 只黑球，b 只白球，从第一袋中取出一球放入第二袋中，然后从第二袋中取出一球放入第三袋中，如此下去，问从最后一个袋中取出一球而为黑球的概率是多少？

8. 飞机有三个不同的部分遭到射击，在第一部分被击中一弹或第二部分被击中两弹，或第三部分被击中三弹时，飞机才能被击落，其命中率与每一部分的面积成正比，设三个部分的面积的百分比为 0.1，0.2，0.7. 若已中两弹，求击落飞机的概率.

*9. 投硬币 n 回，第一回出正面的概率为 c，第二回后每次出现与前一次相同表面的概率为 p，求第 n 回时出正面的概率，并讨论当 $n\to\infty$ 时的情况.

*10. 甲、乙两袋各装一只白球一只黑球，从两袋中各取出一球相交换放入另一袋中，这样进行了若干次. 以 p_n，q_n，r_n 分别记在第 n 次交换后甲袋中将包含两只白球、一只白球一只黑球、两只黑球的概率. 试导出 p_{n+1}，q_{n+1}，r_{n+1} 用 p_n，q_n，r_n 表出的关系式，利用它们求 p_{n+1}，q_{n+1}，r_{n+1} 的表达式，并讨论当 $n\to\infty$ 时的情况.

11. 设一个家庭中有 n 个小孩的概率为

$$p_n=\begin{cases}\alpha p^n, & n\geqslant 1\\ 1-\dfrac{\alpha p}{1-p}, & n=0\end{cases}$$

这里 $0<p<1, 0<\alpha<\dfrac{1-p}{p}$，若认为生一个小孩为男孩或女孩是等可能的，求证一个家庭有 k $(k\geqslant 1)$ 个男孩的概率为 $2\alpha p^k/(2-p)^{k+1}$.

12. 在上题假定下：(1) 已知家庭中至少有一个男孩，求此家庭至少有两个男孩的概率；(2) 已知家庭中没有女孩，求正好有一个男孩的概率.

13. 已知产品中 96% 是合格的，现有一种简化的检查方法，它把真正的合格品确认为合格品的概率为 0.98，而误认废品为合格品的概率为 0.05，求以简化法检查下为合格品的一个产品确实是合格的概率.

14. 炮战中，在距目标 250 米，200 米，150 米处射击的概率分别为 0.1，0.7，0.2，而在各该处射击时命中目标的概率分别为 0.05，0.1，0.2，现在已知目标被击毁，求击毁目标的炮弹是由距目标 250 米处射出的概率.

15. 在通讯渠道中，可传送字符 AAAA，BBBB，CCCC 三者之一，假定传送这三者的概率分别为 0.3，0.4，0.3，由于通道噪声的干扰，正确接收到被传送字母的概率为 0.6，而接收到其他字母的概率是 0.2，假定前后字母是否被歪曲互不影响，若接收到的是 ABCA，问被传送的是 AAAA 的概率.

16. 设 A, B, C 三事件相互独立，求证 $A\cup B$，AB，$A-B$ 皆与 C 独立.

17. 若 A，B，C 相互独立，则 $\overline{A}$，$\overline{B}$，$\overline{C}$ 亦相互独立.

*18. 证明：事件 A_1，A_2，…，A_n 相互独立的充要条件是下列 2^n 个等式成立：

$$P(\hat{A}_1\hat{A}_2\cdots\hat{A}_n)=P(\hat{A}_1)P(\hat{A}_2)\cdots P(\hat{A}_n)$$

其中 $\hat{A}_i$ 取 A_i 或 $\overline{A}_i$.

19. 若 A 与 B 独立，证明 $\{\varnothing, A, \overline{A}, \Omega\}$ 中任何一个事件与 $\{\varnothing, B, \overline{B}, \Omega\}$ 中任何一个事件是相互独立的.

20. 对同一目标进行三次独立射击，第一，二，三次射击的命中概率分别为 0.4，0.5，0.7，试求 (1) 在这三次射击中恰好有一次击中目标的概率；(2) 至少有一次命中目标的概率.

21. 设 A_1，A_2，…，A_n 相互独立，而 $P(A_k)=p_k$，试求：(1) 所有事件全不发生的概率；(2) 诸事件中至少发生其一的概率；(3) 恰好发生其一的概率.

22. 当元件 K 或者元件 K_1 及 K_2 都发生故障时电路断开，元件 K 发生故障的概率等于 0.3，而元件 K_1，K_2 发生故障的概率各为 0.2，求电路断开的概率.

23. 说明"重复独立试验中，小概率事件必然发生"的确切意思.

24. 在第一台车床上制造一级品零件的概率等于 0.7，而在第二台车床上制造此种零件的概率等于 0.8，第一台车床制造了两个零件，第二台制造了三个零件，求所有零件均为一级品的概率.

25. 设实验室器皿中产生甲类细菌与乙类细菌的机会是相同的，若某次发现产生了 $2n$ 个细菌，求 (1) 至少有一个甲类细菌的概率；(2) 甲、乙两类细菌各占其半的概率.

26. 掷硬币出现正面的概率为 p，掷了 n 次，求下列概率：(1) 至少出现一次正面；(2) 至少出现两次正面.

27. 甲、乙、丙三人进行某项比赛，若三人胜每局的概率相等，比赛规定先胜三局者为整场比赛的优胜者，若甲胜了第一、三局，乙胜了第二局，问丙成为整场比赛优胜者的概率是多少？

28. 甲、乙均有 n 个硬币，全部掷完后分别计算掷出的正面数，试求两人掷出的正面数相等的概率.

29. 在伯努利试验中，事件 A 出现的概率为 p，求在 n 次独立试验中事件 A 出现奇数次的概率.

*30. 在伯努利试验中，若 A 出现的概率为 p，试证在出现 m 次 $\bar{A}$ 之前出现 n 次 A 的概率，可由 (2.3.13)，(2.3.14)，(2.3.15) 中的任一形式表出，即它们是相等的.

31. 甲袋中有 $N-1$ 只白球和 1 只黑球，乙袋中有 N 只白球，每次从甲、乙两袋中分别取出一只球并交换放入另一袋中去，这样经过了 n 次，问黑球出现在甲袋中的概率是多少，并讨论 $n\to\infty$ 时的情况.

32. 一个工厂出产的产品中废品率为 0.005，任意取来 1000 件，试计算下面概率：(1) 其中至少有两件废品；(2) 其中不超过 5 件废品；(3) 能以 90% 的概率希望废品件数不超过多少？

33. 某主机有 20 个终端，这些终端被各单位独立操作，使用率各为 0.7，求有 10 个或更多个终端同时操作的概率.

34. 设每次射击打中目标的概率等于 0.001，如果射击 5000 次，试求打中两弹或两弹以上的概率.

35. 某厂长有 7 个顾问，假定每个顾问贡献正确意见的百分比为 0.6，现为某事可行与否而个别征求各顾问意见，并按多数人的意见作出决策，求作出正确决策的概率.

36. 实验室器皿中产生甲、乙两类细菌的机会是相等的，且产生 k 个细菌的概率为

$$p_k=\frac{\lambda^k}{k!}e^{-\lambda},\ k=0,\ 1,\ 2,\ \cdots$$

试求：(1) 产生了甲类细菌但没有乙类细菌的概率；(2) 在已知产生了细菌而且没有甲类细菌的条件下，有 2 个乙类细菌的概率.

37. 假定人在一年 365 日中的任一日出生的概率是一样的，在 50 个人的单位中有两个以上的人生于元旦的概率是多少？

38. 一本 500 页的书，共有 500 个错字，每个字等可能地出现在每一页上，试求在给定的一页上至少有三个错字的概率.

39. 某商店中出售某种商品，据历史记录分析，每月销售量服从泊松分布，参数为 7，问在月初进货时要库存多少件此种商品，才能以 0.999 的概率充分满足顾客的需要.

40. 螺丝钉生产中废品率为 0.015，问一盒应装多少只才能保证每盒中有 100 只以上的好螺丝钉的概率不小于 80%（提示：用泊松逼近，设应装 100 $+k$ 只）.

41. 某疫苗中所含细菌数服从泊松分布，每 1 毫升中平均含有一个细菌，把这种疫苗放入 5 只试管中，每试管放 2 毫升，试求：(1) 5 只试管中都有细菌的概率；(2) 至少有 3 只试管中有细菌的概率.

42. 通过某交叉路口的汽车流可看作泊松过程，若在一分钟内没有车的概率为 0.2，求在 2 分钟内有多于一车的概率.

43. 若每条蚕的产卵数服从泊松分布，参数为 λ，而每个卵变为成虫的概率为 p，且各卵是否变为成虫彼此独立，求每蚕养活 k 只小蚕的概率.

44. 若已知 $t=0$ 时，某分子与另一分子碰撞，又知对任何 $t\geqslant 0$ 和 $\Delta t>0$，若不管该分子在时刻 t 以前是否遭受碰撞，在 $(t,\ t+\Delta t)$ 中遭到碰撞的概率等于 $\lambda\Delta t+o\ (\Delta t)$，试求该分子在时刻 τ 还没有再受到碰撞的概率.

*45. 利用概率论的想法证明下面恒等式：

$$\sum_{k=0}^{N}\binom{N+k}{k}\frac{1}{2^k}=2^N$$

**46. 通过构造适当的概率模型证明：从正整数中随机地选取两数，此两数互素的概率等于$\frac{6}{\pi^2}$.

47. 某车间宣称自己产品的合格率超过 99%，检验人员从该车间的 10000 件产品中抽查了 100 件，发现有两件次品，能否据此断定该车间谎报合格率？

48. 产品验收方案规定：在一批 20 件产品中，抽取其中 4 件，若发现 1 件或 0 件次品，则接受此批产品. 如果一批 20 件产品中含有 5 件次品，若依上述方案验收，试求这批产品被接受的概率.

*49. 系统中每个元件正常工作的概率为 p，有半数元件正常则系统可工作，对什么 p 值，$2k+1$ 个元件的系统比 $2k-1$ 个元件的系统好？

*50. （赠券收集）食品厂把印有水浒 108 将之一的画卡作为赠券装入某种儿童食品袋中，每袋一卡，试求购买 n 袋这种食品而能收齐全套画卡的概率.

第三章　随机变量与分布函数

§1. 随机变量及其分布

一、随机变量的定义

在随机现象中，有很大一部分问题与数值发生关系，例如在产品检验问题中，我们关心的是抽样中出现的废品数；在车间供电问题中我们关心的是某时刻正在工作的车床数；在电话问题中关心的是某段时间中的话务量，它与呼叫的次数及各次呼叫占用交换设备的时间长短有关．此外如测量时的误差，气体分子运动的速度，信号接收机所收到的信号（用电压表示或数字表示）的大小，也都与数值有关．

有些初看起来与数值无关的随机现象，也常常能联系数值来描述．例如在掷硬币问题中，每次出现的结果为正面或反面，与数值没有关系，但是我们能用下面方法使它与数值联系起来，当出现正面时对应数“1”，而出现反面时对应数“0”，为了计算 n 次投掷中出现的正面数就只须计算其中“1”出现的次数了．

一般地，如果 A 为某个随机事件，则一定可以通过如下示性函数使它与数值发生联系：

$$1_A=\begin{cases}1,\ \text{如果 } A \text{ 发生}\\0,\ \text{如果 } A \text{ 不发生}\end{cases}$$

这些例子中，试验的结果能用一个数 ξ 来表示，这个数 ξ 是随着试验的结果的不同而变化的，也即它是样本点的一个函数，这种量以后称为**随机变量**（random variable）．本书中将主要用希腊字母 ξ，η，ζ，…来表示随机变量．下面我们就来考虑应当如何给

这种量以严格的数学定义.

正如对随机事件一样，我们所关心的不仅是试验会出现什么结果，更重要的是要知道这些结果将以怎样的概率出现，也即对随机变量，我们不但要知道它取什么数值，而且要知道它取这些数值的概率.

从随机现象的可能出现的结果来看，随机变量至少有两种不同的类型. 一种是试验结果 ξ 所可能取的值为有限个或至多可列个，我们能把其可能结果一一列举出来，这种类型的随机变量称为**离散型随机变量**. 在日常生活中经常碰到离散型随机变量，例如废品数、电话呼叫数等等. 前面讨论过的随机现象大部分都能用离散型随机变量来描述. 例如古典概型中只有有限个可能结果，若对应于每一个结果用一个数值来表征，则得到一个离散型随机变量. 又如在 n 次伯努利试验中，若以 μ 记事件 A 出现的次数，则 μ 可取值 0，1，2，…，n. 在呼叫流的研究中，若以 $\xi(t)$ 记 $[0, t]$ 中来到的呼叫数，则 $\xi(t)$ 可取值 0，1，2，…这些都是离散型随机变量.

从上章的讨论可以看到，要描述这类随机变量并不难，以 n 次伯努利试验中事件 A 出现的次数 μ 为例，我们知道它是样本点 ω 的函数，也就是说严格来讲，应写作 $\mu(\omega)$，其中 $\omega\in\Omega$，它取的值是 0，1，2，…，n，并且知道 $\mu(\omega)$ 取这些值的概率为[①]

$$P\{\mu(\omega)=k\}=\binom{n}{k}p^kq^{n-k},\ k=0,1,2,\cdots,n$$

---

① 下面出现记号 $\{\mu(\omega)=k\}$，对于不熟悉这类记号的读者，我们作如下说明. $\{\mu(\omega)=k\}$ 是 $\{\omega: \mu(\omega)=k\}$ 的简写，它表示具有如下性质的样本点 ω 的集合：在其上 μ 取固定值 k. 一般地，对于从样本空间 Ω 到数直线 $\mathbf{R}^1$ 上的单值映射 $\xi(\omega)$，若 A 是 $\mathbf{R}^1$ 的某一子集，常以 $\{\xi(\omega)\in A\}$ 作为 $\{\omega: \xi(\omega)\in A\}$ 的简写，用来表示使 $\xi(\omega)$ 之值属于 A 的那些样本点 ω 的集合，有时还进一步简记作 $\xi^{-1}(A)$；特别地，$\{\xi(\omega)<x\}$ 即为 $\{\omega: \xi(\omega)\in(-\infty, x)\}$，这里 x 是某一实数.

这里我们已经知道了 $\mu(\omega)$ 取什么值，以及以什么概率取这些值.

一般地，对于定义在样本空间 Ω 上的离散型随机变量 $\xi(\omega)$ 只要能指出它取的值 x_1，x_2，…，x_n，…以及它取这些值的概率 $P\{\xi(\omega)=x_i\}$，$i=1, 2, \cdots, n, \cdots$就满足了我们的要求. 显然要做到这一点，必须要求 $\{\xi(\omega)=x_i\}$ 有概率. 因为我们只对事件域 $\mathscr{F}$ 中的集合定义概率，所以必须有 $\{\omega: \xi(\omega)=x_i\}\in\mathscr{F}$.

与离散型随机变量不同，一些随机现象所出现的试验结果 ξ 不止取可列个值，例如测量误差、分子运动速度、候车时的等待时间、降水量、风速、洪峰值等等皆是. 这时用来描述试验结果的随机变量还是样本点 ω 的函数：严格写应是 $\xi(\omega)$，其中 $\omega\in\Omega$，但是这随机变量能取某个区间 $[c, d]$ 或 $(-\infty, \infty)$ 的一切值.

假如想用描述离散型随机变量的方法（简单地罗列所取的值及相应的概率）来描述这后一类随机变量，则会碰到很大的困难. 一来是这类随机变量所取的值不能一一列出；二来，我们下面将会看到，取连续值的随机变量，它取某个特定值的概率常常是 0，因此用这种描述方法根本不行.

对于取连续值的随机变量我们所关心的也并不是它取某个特定值的概率. 例如在测量误差的讨论中，我们感兴趣的是测量误差小于某个数的概率；在降雨问题中，我们重视的是雨量在某一个量级，例如在 100 毫米到 120 毫米之间的概率. 总之，对于取连续值的随机变量 $\xi(\omega)$，我们感兴趣的是 $\xi(\omega)$ 取值于某个区间 (a, b) 的概率，或取值于若干个这种区间的概率. 因此应当要求 $\{a\leqslant\xi(\omega)<b\}$或$\{\xi(\omega)<b\}$或一般地$\{\xi(\omega)\in A\}$（其中 A 是由区间经并、交等运算而得到的直线上的某一个点集）有概率可言，既然只对概率空间 $(\Omega, \mathscr{F}, P)$ 的事件域 $\mathscr{F}$ 中的集合才定义概率，因此我们自然要求上述集合属于 $\mathscr{F}$，即都是事件.

通过上面讨论可以看出，为了使我们感兴趣的概率计算得以进行，我们应对 $\xi(\omega)$ 加上一定的限制，主要是要求 $\{\omega: \xi(\omega)\in B\}$ 应是事件. 在离散型随机变量的场合，B 是直线上的某一个点；在

取连续值场合，B 是直线上由区间经并、交等运算而得到的某一点集．在概率计算中有时要考虑可列运算，因此较方便的是取 B 为直线上博雷尔点集．

为此引进如下定义：

定义 3.1.1 设 $\xi(\omega)$ 是定义于概率空间 $(\Omega, \mathscr{F}, P)$ 上的单值实函数，如果对于直线上任一博雷尔点集 B，有

$$\{\omega: \xi(\omega) \in B\} \in \mathscr{F} \tag{3.1.1}$$

则称 $\xi(\omega)$ 为**随机变量**（图 15），而 $P\{\xi(\omega) \in B\}$ 称为随机变量 $\xi(\omega)$ 的**概率分布**．

特别地，若取 $B=(-\infty, x)$，则有

$$\{\omega: \xi(\omega) < x\} \in \mathscr{F} \tag{3.1.2}$$

因此 $P\{\xi(\omega) < x\}$ 有定义．注意到

$$P\{a \leqslant \xi(\omega) < b\} = P\{\xi(\omega) < b\} - P\{\xi(\omega) < a\} \tag{3.1.3}$$

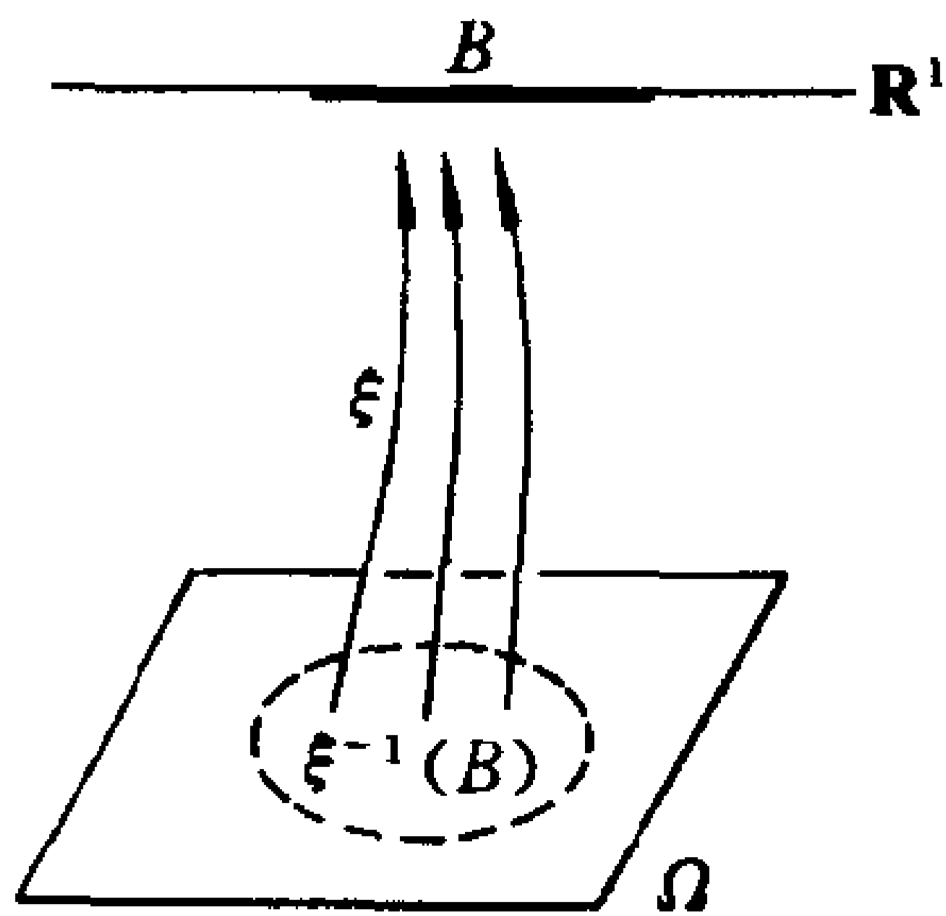


图 15 随机变量

所以只要对一切实数 x 给出概率 $P\{\xi(\omega) < x\}$，就能算出 $\xi(\omega)$ 落入某个区间 $[a, b)$ 的概率，再利用概率的性质还可以算出 $\xi(\omega)$ 属于某些相当复杂的直线点集（譬如可列个不相交的左闭右开区间之和）的概率．

定义 3.1.2 称

$$F(x) = P\{\xi(\omega) < x\}, \quad -\infty < x < \infty \tag{3.1.4}$$

为随机变量 $\xi(\omega)$ 的**分布函数**（distribution function）．

为书写方便起见，通常把"随机变量 $\xi(\omega)$ 服从分布函数 $F(x)$"简记作 $\xi(\omega) \sim F(x)$．

由（3.1.3）立刻得到

$$P\{a \leqslant \xi(\omega) < b\} = F(b) - F(a) \tag{3.1.5}$$

在上面讨论中，我们根据描述随机变量的需要给出了随机变

量与分布函数的定义. 按定义，随机变量是样本点的函数，因此在试验前我们只能知道它可能取哪些值，而不能确知它将取何值，这就是随机性；但是到了试验之后，它的取值也就明确了。为了计算概率，必须要求随机变量具有可测性（3.1.1），而分布函数的引进则把对于随机变量的概率计算化为对分布函数的数值运算. 这样一来，我们已经在科尔莫戈罗夫的公理化结构中给随机变量予严格的定义，同时又为对它的研究准备了方便的分析工具.

应该指出，最后之所以采用这种定义，还有数学理论上的需要，有些进一步的事实，例如若（3.1.2）成立则（3.1.1）也成立，由分布函数可以唯一决定概率分布等等是可以通过测度论的方法证明的，这些已超出本课程范围，我们将不予讨论.

概率论是研究大量随机现象中的数量规律的数学分支，研究随机变量和分布函数是它的重要任务，而且概率论中所研究的也大都局限于能用随机变量来描述的随机现象. 前面已指出，随机事件的研究可通过示性函数转化为对随机变量的研究，随机变量是取数值的，因此可以对它进行各种数学运算，研究起来就很方便.

二、分布函数的性质

我们先把分布函数的最基本的性质汇集于下列定理中.

定理 3.1.1 分布函数 $F(x)$ 具有下列性质：

(i) 单调性：若 $a<b$，则 $F(a)\leqslant F(b)$；

(ii) $\lim\limits_{x\to-\infty} F(x)=0,\ \lim\limits_{x\to+\infty} F(x)=1$； (3.1.6)

(iii) 左连续性：$F(x-0)=F(x)$.

[证明] (i) $F(b)-F(a)=P\{a\leqslant\xi<b\}\geqslant 0$.

$$\text{(ii)}\ P\{-\infty<\xi<+\infty\}=\sum_{n=-\infty}^{\infty}P\{n\leqslant\xi<n+1\}$$

$$=\sum_{n=-\infty}^{\infty}[F(n+1)-F(n)]$$

$$=\lim_{n\to+\infty}F(n)-\lim_{m\to-\infty}F(m)=1$$

由于 $F(x)$ 的单调性，$\lim\limits_{x\to-\infty}F(x)=\lim\limits_{m\to-\infty}F(m)$，$\lim\limits_{x\to+\infty}F(x)=\lim\limits_{n\to+\infty}F(n)$ 存在.

因为 $0\leqslant F(x)\leqslant 1$，故

$$\lim_{x\to-\infty}F(x)=0,\ \lim_{x\to+\infty}F(x)=1$$

今后为书写方便起见，将记

$$F(-\infty)=\lim_{x\to-\infty}F(x),F(+\infty)=\lim_{x\to+\infty}F(x)\qquad(3.1.7)$$

(iii) 由于 $F(x)$ 是单调函数，只须证明对于一列单调上升的数列 $x_0<x_1<x_2<\cdots<x_n<\cdots$，$x_n\to x$ 成立 $\lim\limits_{n\to\infty}F(x_n)=F(x)$ 即可.

因为

$$F(x)-F(x_0)=P\{x_0\leqslant\xi<x\}=\sum_{n=1}^{\infty}[F(x_n)-F(x_{n-1})]$$

$$=\lim_{n\to\infty}F(x_n)-F(x_0)$$

所以

$$F(x-0)=\lim_{n\to\infty}F(x_n)=F(x)$$

可以看出分布函数的这三个基本性质，正好对应于概率的三个基本性质.

有了分布函数，关于随机变量 $\xi(\omega)$ 的许多概率都能方便算出，例如

$$P\{\xi(\omega)=a\}=F(a+0)-F(a)$$

$$P\{\xi(\omega)\leqslant a\}=F(a+0)$$

$$P\{\xi(\omega)\geqslant a\}=1-F(a)$$

$$P\{\xi(\omega)>a\}=1-F(a+0)$$

综上所述，分布函数是一种分析性质良好的函数，便于处理；而给定了分布函数就能算出各种事件的概率。因此引进分布函数之后，许多概率论问题便简化或归结为函数的运算，这样就能利用数学分析的许多结果. 这也是引进随机变量的好处之一.

可以证明满足上述定理中三个性质的函数必是某随机变量的分布函数，关于这点的讨论将在第3节进行.

对于随机变量及其概率分布的研究，若按随机变量的不同类型分别讨论是有好处的，下面我们将照此线索进行.

三、离散型随机变量

设 $\{x_i\}$ 为离散型随机变量 ξ 的所有可能值；而 $p(x_i)$是 ξ 取 x_i 的概率，即

$$P\{\xi=x_i\}=p(x_i), i=1,2,3,\cdots \tag{3.1.8}$$

$\{p(x_i), i=1,2,3,\cdots\}$称为随机变量 ξ 的**概率分布**，它应满足下面关系：

$$p(x_i)\geqslant 0, \quad i=1,2,3,\cdots \tag{3.1.9}$$

$$\sum_{i=1}^{\infty} p(x_i) = 1 \tag{3.1.10}$$

当给定了 $\{x_i, i=1, 2, \cdots\}$ 及 $\{p(x_i), i=1, 2, \cdots\}$，我们就能很好地描述随机变量 ξ，因为我们已经知道了它取什么值，以及以什么概率取这些值，而这正是我们对随机变量所关心的问题.

常用下面方法来表出离散型随机变量 ξ 的概率分布，

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p(x_1) & p(x_2) & \cdots & p(x_n) & \cdots \end{pmatrix} \tag{3.1.11}$$

(3.1.11) 称为随机变量 ξ 的**分布列**，由分布列能一目了然地看出随机变量 ξ 的取值范围及取这些值的概率.

有了分布列，可以通过下式求得分布函数

$$F(x) = P\{\xi < x\} = \sum_{x_k < x} p(x_k) \tag{3.1.12}$$

显然这时 $F(x)$是一个跳跃函数，它在每个 x_k 处有跳跃度$p(x_k)$. 当然，由 $F(x)$也可唯一决定 x_k 及 $p(x_k)$，因此用分布列或分布函数都能描述离散型随机变量.

图 16(a) 中显示一个离散型随机变量的概率分布，而图 16

(b) 中则是它对应的分布函数.

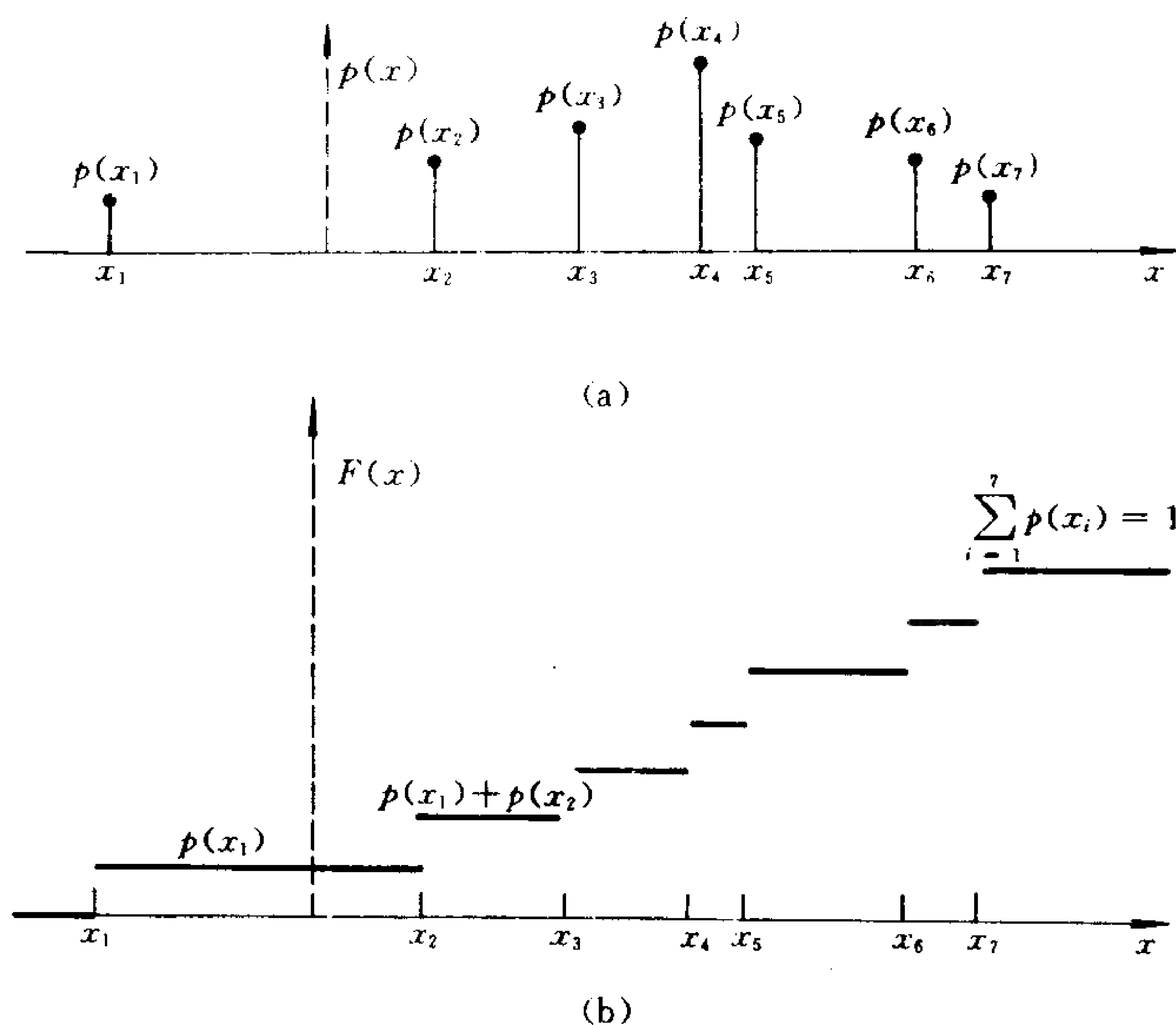


图 16 离散型概率分布与分布函数

下面看一些离散型随机变量及其概率分布的例子. 这些分布大都在上两章引入过.

[退化分布] 若随机变量 α 只取常数值 c，即

$$P\{\alpha=c\}=1$$

这时分布函数为

$$I(x-c)=\begin{cases}0, x\leqslant c \\ 1, x>c\end{cases} \tag{3.1.13}$$

可以说，这样的 α 并不随机，但有时我们宁愿把它看作随机变量的退化情况更为方便，因此称之为**退化分布**，又称**单点分布**.

[伯努利分布] 在一次试验中，事件 A 出现的概率为 p，不出现的概率为 $q=1-p$，若以 β 记事件 A 出现的次数，则 β 仅取 0，1 两值，相应的概率分布为

$$b_k=P\{\beta=k\}=p^k q^{1-k}, k=0,1 \tag{3.1.14}$$

这个分布称为**伯努利分布**，亦称**两点分布**.

[二项分布] 在 n 重伯努利试验中，若以 μ 记事件 A 出现的次数，则它是一个随机变量，μ 可能取的值为 0，1，2，…，n，其对应的概率由**二项分布**给出：

$$b(k;n,p)=P\{\mu=k\}=\binom{n}{k}p^k q^{n-k}$$

$$k=0,\ 1,\ 2,\ \cdots,\ n \qquad (3.1.15)$$

简记作 $\mu\sim B\ (n,\ p)$.

关于二项分布及其计算已在上章作过详细讨论，这里不再重复. 至于为什么称为分布，到了此处自然十分明白.

顺便指出，伯努利分布可以看作 $n=1$ 时的二项分布，这时相应于一次独立试验的场合.

[超几何分布] 对某批 N 件产品进行不放回抽样检查，若这批产品中有 M 件次品，现从整批产品中随机抽出 n 件产品，则在这 n 件产品中出现的次品数 ν 是随机变量，它取值 0，1，2，…，n，其概率分布为**超几何分布**.

$$h_k=P\{\nu=k\}=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}},\quad \begin{matrix}0\leqslant k\leqslant n\leqslant N\\ k\leqslant M\end{matrix} \qquad (3.1.16)$$

这个分布在第一章中已经出现，我们还证明过当 N 很大而 n 较小时，它可用二项分布来近似.

[泊松分布] 若随机变量 ξ 可取一切非负整数值，且

$$P\{\xi=k\}=\frac{\lambda^k}{k!}\mathrm{e}^{-\lambda},\ k=0,\ 1,\ 2,\ \cdots \qquad (3.1.17)$$

其中 $\lambda>0$，则称 ξ 服从**泊松分布**. 简记作 $\xi\sim P\ (\lambda)$.

泊松分布是概率论中非常重要的一个分布，在第二章§4 中已对它进行过初步讨论，以后我们还会经常提到它.

[几何分布] 在事件 A 发生的概率为 p 的伯努利试验中，若

以 η 记 A 首次出现时的试验次数，则 η 为随机变量，它可能取的值为 1，2，3，…其概率分布为**几何分布**：

$$g(k,p)=P\{\eta=k\}=q^{k-1}p, \quad k=1,2,\cdots \tag{3.1.18}$$

作为一种等待分布，我们在前两章中已多次碰到过几何分布，这也说明了几何分布是一种常见的概率分布. 几何分布在概率论中的重要性，还在于它具有下面特殊的性质.

几何分布的无记忆性 如上所述，在伯努利试验中，等待首次成功的时间 η 服从几何分布 (3.1.18). 现在假定已知在前 m 次试验中没有出现成功，那么为了达到首次成功所再需要的等待时间 η'，其概率分布为

$$\begin{aligned}P\{\eta'=k\}&=P\{\eta=m+k\,|\,\eta>m\}\\&=\frac{P\{\eta=m+k\}}{P\{\eta>m\}}=\frac{q^{m+k-1}p}{q^m}\\&=q^{k-1}p, \qquad k=1,2,\cdots\end{aligned}$$

还是服从几何分布 (3.1.18)，与前面的失败次数 m 无关，形象化地说，就是把过去的经历完全忘记了. 因此无记忆性是几何分布所具有的一个有趣的性质.

更加有趣的是，在离散型分布中，只有几何分布才具有这样一种特殊的性质. 下面我们来严格叙述并证明这个事实.

若 η 是取正整数值的随机变量，并且，在已知 $\eta>k$ 的条件下，$\eta=k+1$ 的概率与 k 无关，那么 η 服从几何分布.

这个论断的证明如下：以 p 记上述条件概率，并记 $q_k=P\{\eta>k\}$ 及 $p_k=P\{\eta=k\}$，那么 $p_{k+1}=q_k-q_{k+1}$，而且在已知 $\eta>k$ 的条件下，$\eta=k+1$ 的条件概率为 $\frac{p_{k+1}}{q_k}$，因此

$$\frac{p_{k+1}}{q_k}=p$$

即

$$\frac{q_{k+1}}{q_k}=1-p$$

注意到 $q_0=1$，那么 $q_k=(1-p)^k$，因此

$$p_k=(1-p)^{k-1}p,\quad k=1,2,\cdots$$

这正是几何分布（3.1.18）.

［巴斯卡分布］ 在上例的伯努利试验中，若以 ζ 记第 r 次成功出现时的试验次数，则 ζ 是随机变量，取值 r，$r+1$，…其概率分布为**巴斯卡分布**：

$$P\{\zeta=k\}=\binom{k-1}{r-1}p^r q^{k-r},\quad k=r,r+1,\cdots \qquad (3.1.19)$$

这分布在上章 § 3 出现过．显然当 $r=1$ 时，即为几何分布.

另一方面若以 η_i 记从第 $i-1$ 次成功之后的第一次试验算起至第 i 次成功为止共进行的试验次数，则 η_i 服从几何分布（3.1.18），而且

$$\zeta=\eta_1+\cdots+\eta_r \qquad (3.1.20)$$

若记 $\tilde{\zeta}=\zeta-r$，则 $\tilde{\zeta}$ 表示为等待第 r 次成功所经历过的失败次数，那么，

$$P\{\tilde{\zeta}=l\}=P\{\zeta=r+l\}=\binom{r+l-1}{r-1}p^r q^{r+l-r}$$

$$=\binom{r+l-1}{l}p^r q^l=\binom{-r}{l}p^r(-q)^l,l=0,1,2,\cdots$$

$$(3.1.21)$$

显然，ζ 与 $\tilde{\zeta}$ 只是计数方式的不同，一个记全部试验数（包括成功次数与失败次数），另一个则只记失败的次数，它们描述的是同样的随机模型.

这样定义的概率满足非负性及

$$\sum_{l=0}^{\infty}\binom{-r}{l}p^r(-q)^l=p^r(1-q)^{-r}=1$$

因此（3.1.21）也表示一个概率分布，不少书上把这个分布作为巴斯卡分布的定义.

特别地，取 $r=1$，考察为等待第 1 次成功所经历过的失败次数 $\tilde{\eta}$，那么，

$$P\{\tilde{\eta}=l\}=q^l p,\ l=0,\ 1,\ 2,\ \cdots \qquad (3.1.22)$$

这也是一个分布，表示等待首次成功所经历过的失败数，也称为几何分布.

有趣的是，等待第 r 次成功所经历过的失败次数也可作如下分解：

$$\tilde{\xi}=\tilde{\eta}_1+\cdots+\tilde{\eta}_r$$

其中 $\tilde{\eta}_i$ 也服从分布（3.1.22）.

总之，几何分布与巴斯卡分布都有两种表达式，本书采用前者，即（3.1.18）及（3.1.19），阅读参考书时务请注意.

对巴斯卡分布，可以略加推广，即去掉 r 是正整数的限制，这便得到

[负二项分布]　对于任意实数 $r>0$，称

$$Nb(l;r,p)=\binom{-r}{l}p^r(-q)^l,\quad l=0,1,2,\cdots \qquad (3.1.23)$$

为**负二项分布**.

从前面的推导过程中很容易看出这的确是一个概率分布，它的取名也很自然.

负二项分布作为某些离散随机现象的数学模型，正在逐步引起人们的注意.

四、连续型随机变量

除了离散型随机变量之外，还有一类重要的随机变量——**连续型随机变量**. 这种随机变量 ξ 可取某个区间 $[c,\ d]$ 或 $(-\infty,\ \infty)$ 中的一切值，而且其分布函数 $F(x)$是绝对连续函数，即存在可积函数 $p(x)$，使

$$F(x)=\int_{-\infty}^{x}p(y)\mathrm{d}y \qquad (3.1.24)$$

称 $p(x)$ 为 ξ 的（分布）**密度函数**（dlnsity function）.

显然

$$p(x)=F'(x) \tag{3.1.25}$$

由分布函数的性质可知对 $p(x)$ 应有

$$p(x)\geqslant 0 \tag{3.1.26}$$

$$\int_{-\infty}^{\infty} p(x)\mathrm{d}x = 1 \tag{3.1.27}$$

反之，对于定义在 $(-\infty, \infty)$ 上的可积函数 $p(x)$，若它满足（3.1.26）及（3.1.27），则由（3.1.24）定义的函数 $F(x)$ 是一个分布函数，即它有分布函数所必须具备的三个性质.

顺便指出，由于在若干个点上甚至一个零测集上改变被积函数 $p(x)$ 的值，都不影响积分 $F(x)$ 之值，因此，关于 $p(x)$ 的论断通常都是在“几乎处处”的意义上成立，今后不再一一提及.

由（3.1.5）立刻得到

$$P\{a \leqslant \xi < b\} = F(b) - F(a) = \int_a^b p(x)\mathrm{d}x \tag{3.1.28}$$

因此给定密度函数,便可以算出随机变量落入某一个区间的概率.

下面计算 $P\{\xi=c\}$，因为

$$P\{\xi = c\} \leqslant P\{c \leqslant \xi < c + h\} = \int_c^{c+h} p(x)\mathrm{d}x$$

故

$$0 \leqslant P\{\xi = c\} \leqslant \lim_{h\to 0}\int_c^{c+h} p(x)\mathrm{d}x = 0$$

因此

$$P\{\xi=c\}=0 \tag{3.1.29}$$

即连续型随机变量取个别值的概率为 0，这与离散型随机变量截然不同. 因此用列举连续型随机变量取某个值的概率来描述这种随机变量不但做不到，而且也毫无意义.

此外，上述结果还表明，一个事件的概率等于零，这事件并不一定是不可能事件；同样地，一个事件的概率等于 1，这事件也不一定是必然事件.

由于在 $p(x)$的连续点 x 处，

$$p(x)\Delta x \approx \int_{x}^{x+\Delta x} p(y)\mathrm{d}y = F(x+\Delta x) - F(x) \tag{3.1.30}$$

因此密度函数 $p(x)$ 的数值反映了随机变量取 x 的邻近的值的概率的大小. 所以用密度函数描述连续型随机变量的概率分布在某种意义上与离散型时用分布列描述，又有相似之处.

下面举一些常见的连续型分布的例子.

[均匀分布] 若 a，b 为有限数，由下列密度函数定义的分布称为 $[a, b]$ 上**均匀分布**：

$$p(x) = \begin{cases} \dfrac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & x < a \text{ 或 } x > b \end{cases} \tag{3.1.31}$$

相应的分布函数为

$$F(x) = \begin{cases} 0, & x \leqslant a \\ \dfrac{x-a}{b-a}, & a < x \leqslant b \\ 1, & x > b \end{cases} \tag{3.1.32}$$

$[a, b]$ 上均匀分布有时简记作 $U[a, b]$.

若随机变量 ξ 服从 $[a, b]$ 上均匀分布，则 ξ 在 $[a, b]$ 中取值落在某一区域内的概率与这个区域的测度成正比. 粗略地讲就是，ξ 取 $[a, b]$ 中任一点的可能性一样. 当然也可以反过来看，均匀分布正是把这种直观的讲法严格化.

图 17 画出了 $[a, b]$ 上均匀分布的密度函数及分布函数.

[例 1] 定点计算中的舍入误差可以作为常见的均匀分布随机变量的例子. 假如我们在运算中，数据都只保留到小数点后第五位，而小数点第五位以后的数字按四舍五入处理. 若以 x 表示真值，以 $\hat{x}$ 表示舍入后的值，则误差 $\varepsilon = x - \hat{x}$ 一般假定为 $[-0.5\times10^{-5}, 0.5\times10^{-5}]$ 上均匀分布的随机变量，有了这个假定，就能对经过大量运算后的数据进行误差分析，这种误差分析

在用数字计算机解题时是很必要的，因为数字计算机字长总是有限的.

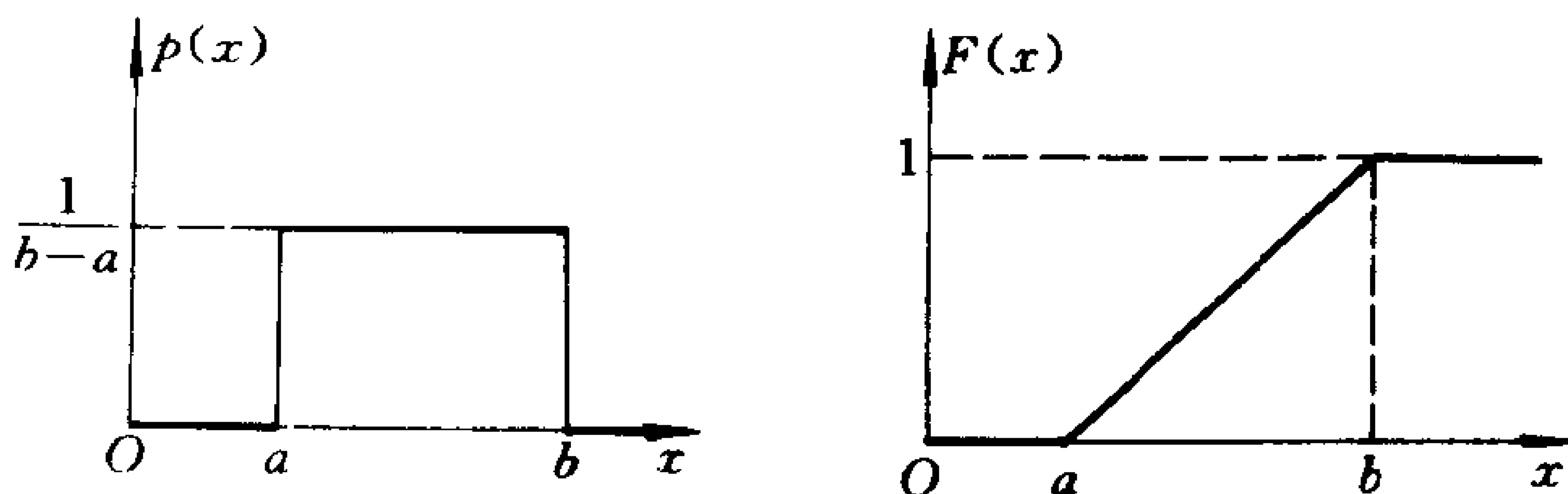


图 17　均匀分布的密度函数与分布函数

[正态分布]　密度函数为

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}},-\infty<x<\infty \qquad (3.1.33)$$

其中 $\sigma>0$，a 与 σ 均为常数，相应的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(y-a)^2}{2\sigma^2}}dy,\ -\infty<x<\infty \qquad (3.1.34)$$

这分布称为**正态分布**（normal distribution），简记为 $N(a,\sigma^2)$.

特别当 $a=0$，$\sigma=1$，这时分布称为**标准正态分布**，记为 $N(0,1)$，相应的分布密度函数及分布函数分别记为 $\varphi(x)$ 及 $\Phi(x)$，见图 18 及图 19.

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}},\ -\infty<x<\infty \qquad (3.1.35)$$

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{y^2}{2}}dy,\ -\infty<x<\infty \qquad (3.1.36)$$

习惯上把服从正态分布的随机变量称为**正态变量**. 可以验证，若随机变量 ξ 服从正态分布 $N(a,\sigma^2)$，简记作 $\xi\sim N(a,\sigma^2)$，则随机变量 $\eta=\frac{\xi-a}{\sigma}$ 服从 $N(0,1)$.

为了说明（3.1.33）确实定义了一个密度函数，需要验证关系式（3.1.26）及（3.1.27），显然

$$p(x)>0$$

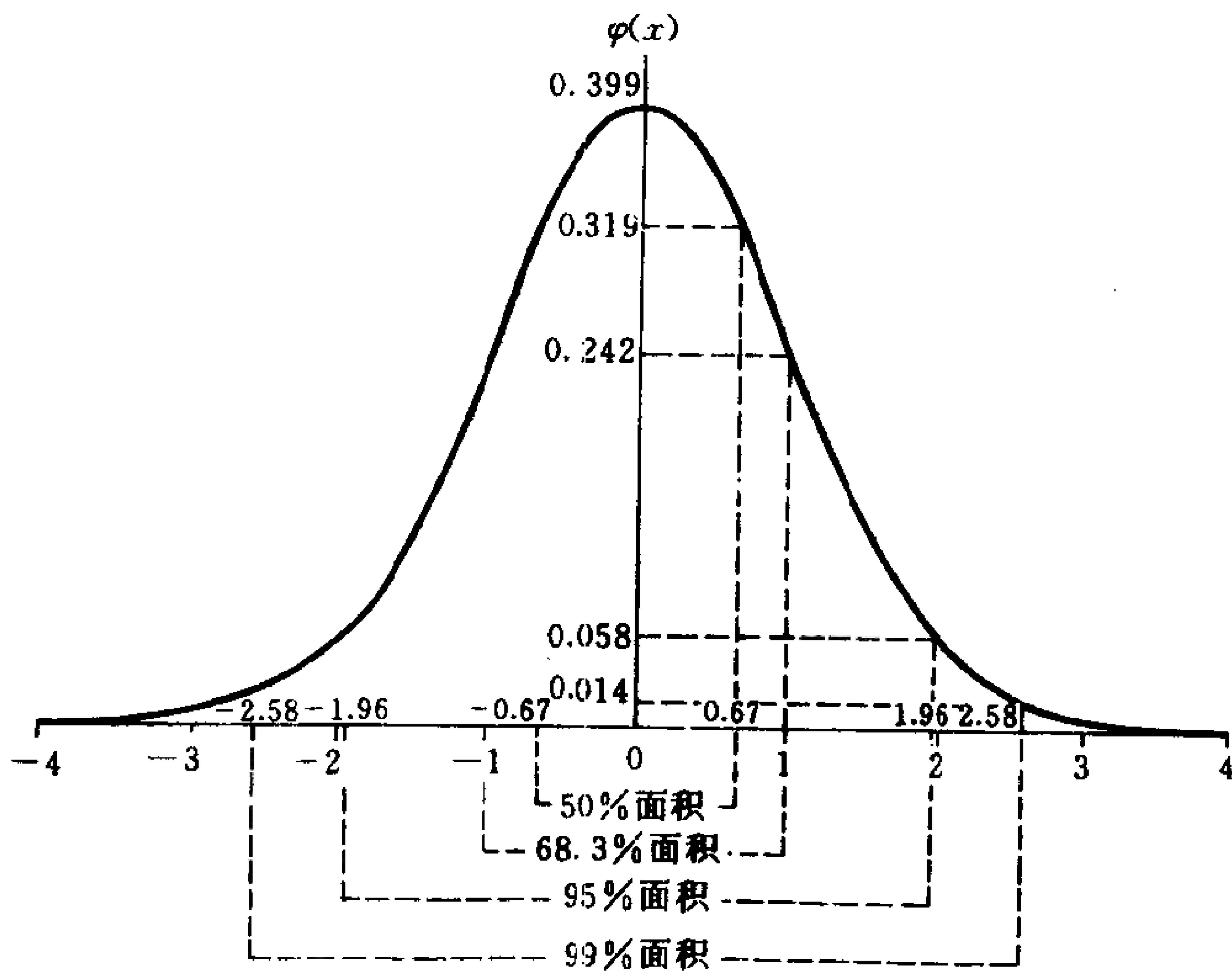


图 18　正态密度函数 $\varphi(x)$

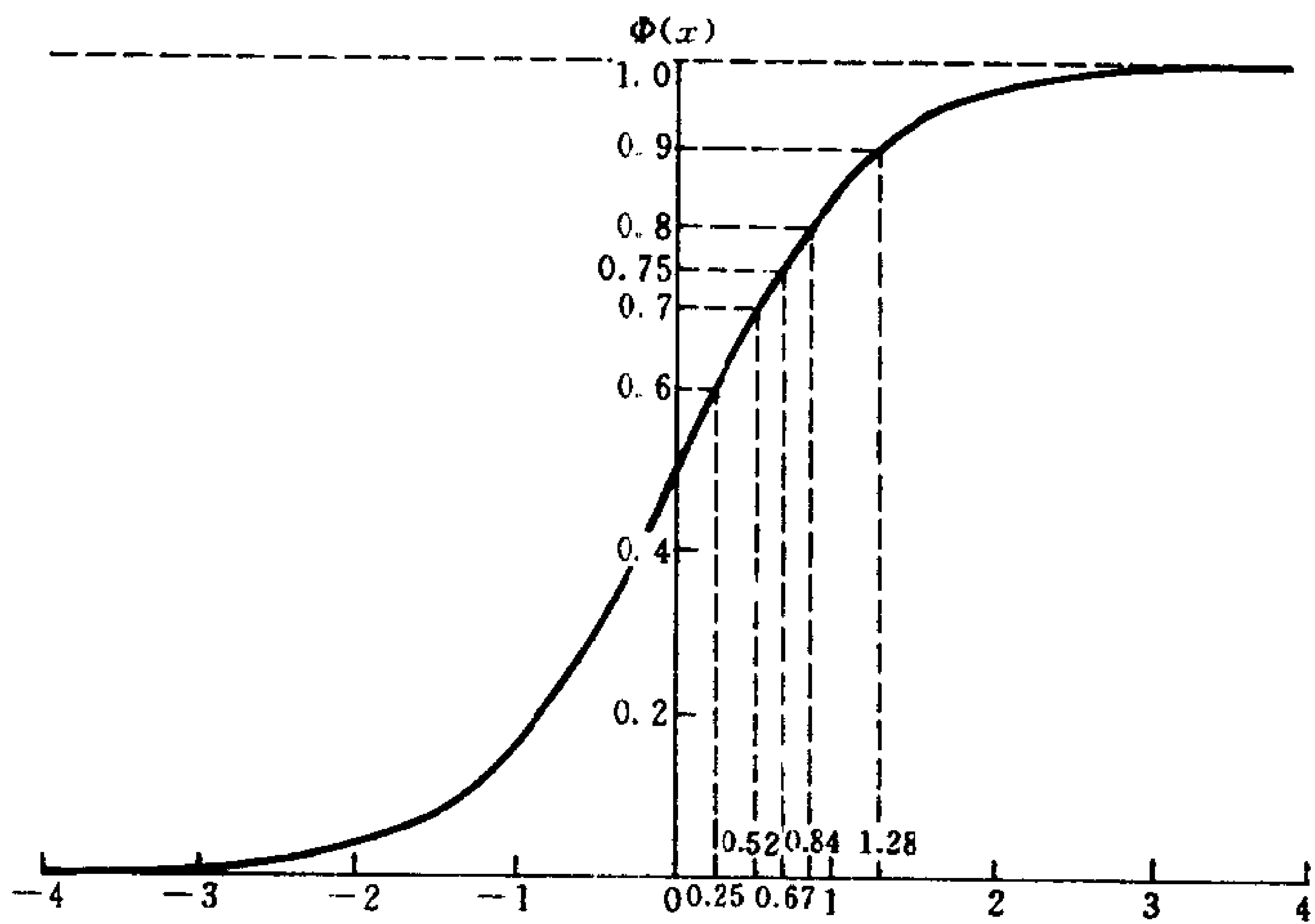


图 19　正态分布函数 $\Phi(x)$

因此剩下的是验证（3.1.27）. 令$\frac{x-a}{\sigma}=y$，则

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy$$

但是

$$\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx\right)\left(\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy\right)$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dxdy$$

变换到极坐标，令 $x=r\cos\varphi$，$y=r\sin\varphi$，则

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{x^2+y^2}{2}}dxdy$$

$$=\frac{1}{2\pi}\int_{0}^{\infty}\int_{0}^{2\pi}e^{-\frac{r^2}{2}}rdrd\varphi$$

$$=\int_{0}^{\infty}re^{-\frac{r^2}{2}}dr=1$$

由$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx$的非负性知

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy=1$$

从而完成了（3.1.27）的验证.

正态分布是概率论中最重要的分布. 一方面，正态分布是自然界最常见的一种分布，例如测量的误差；炮弹弹落点的分布；人的生理特征的尺寸：身高，体重等；农作物的收获量；工厂产品的尺寸：直径，长度，宽度，高度；……都近似服从正态分布. 一般说来，若影响某一数量指标的随机因素很多，而每个因素所起的作用不太大，则这个指标服从正态分布，这点可以利用概率论的极限定理来加以证明. 另一方面，正态分布具有许多良好的性质，许多分布可用正态分布来近似，另外一些分布又可以通过正态分布来导出，因此在理论研究中，正态分布十分重要.

一般正态分布密度函数 $p(x)$的图形见图 20，通过图形或(3.1.33)不难看出：$p(x)$在 $x=a$ 处达到极大，整个图形关于$x=a$对称. 当 σ 不同时，$p(x)$的形状也不同，σ 越小，分布越集中在 $x=a$ 附近；当 σ 越大时，分布就越平坦.

由于正态分布在概率计算中的重要性，已编造了各种各样的正态分布表，本书的附表中就有 $a=0$，$\sigma=1$ 时的正态分布密度函数 $\varphi(x)$和正态分布函数 $\Phi(x)$的表. 由于 $\varphi(-x)=\varphi(x)$及$\Phi(-x)=1-\Phi(x)$，所以表中只对正的 x 给出 $\varphi(x)$ 及$\Phi(x)$的数值. 一般 $N(a,\sigma^2)$ 的分布函数值可由变换而得

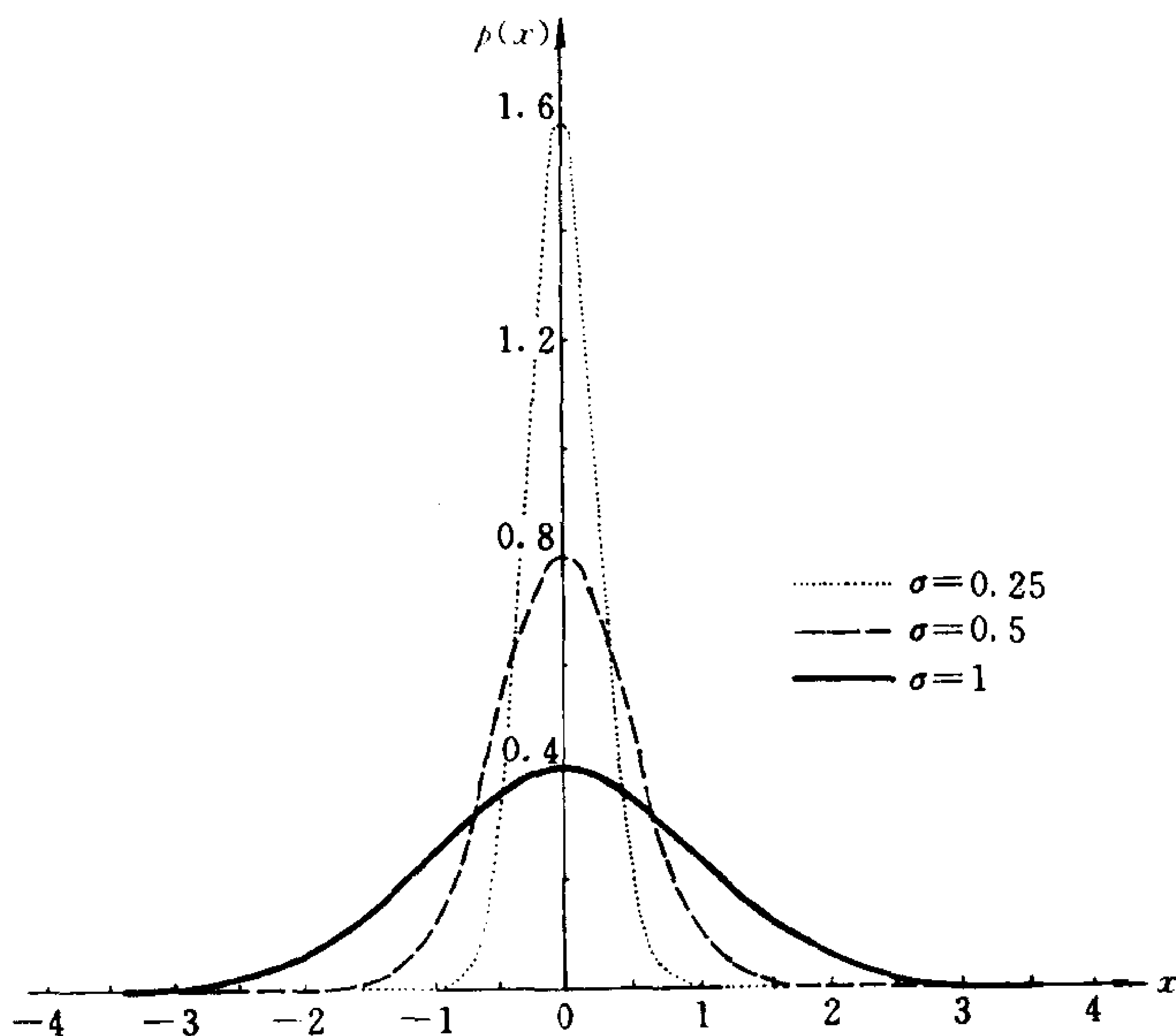


图 20　$a=0$ 且具有不同的 σ^2 的正态密度曲线

$$F(x)=\Phi\left(\frac{x-a}{\sigma}\right) \qquad (3.1.37)$$

从表中可以看出，若 ξ 服从 $N(a,\sigma^2)$，则

$$P\{|\xi-a|<\sigma\}\approx 68.27\% \tag{3.1.38}$$

$$P\{|\xi-a|<2\sigma\}\approx 95.45\% \tag{3.1.39}$$

$$P\{|\xi-a|<3\sigma\}\approx 99.73\% \tag{3.1.40}$$

因此可以说，在一次试验里，ξ 几乎总是落在 $(a-3\sigma, a+3\sigma)$ 之中.

下例说明正态分布与实测数据符合得很好.

[例 2] 某手表厂曾对其生产的某个零件的重量收集了大量资料，对测量得的 3805 个数据，按不同重量加以分组，并记录了不同范围内零件的个数（频数），计算了它们的频率，结果如下表所示，它们与 $a=56.94$，$\sigma=8.2$ 的正态分布符合得相当好，图 21 表示了两者符合的情况.

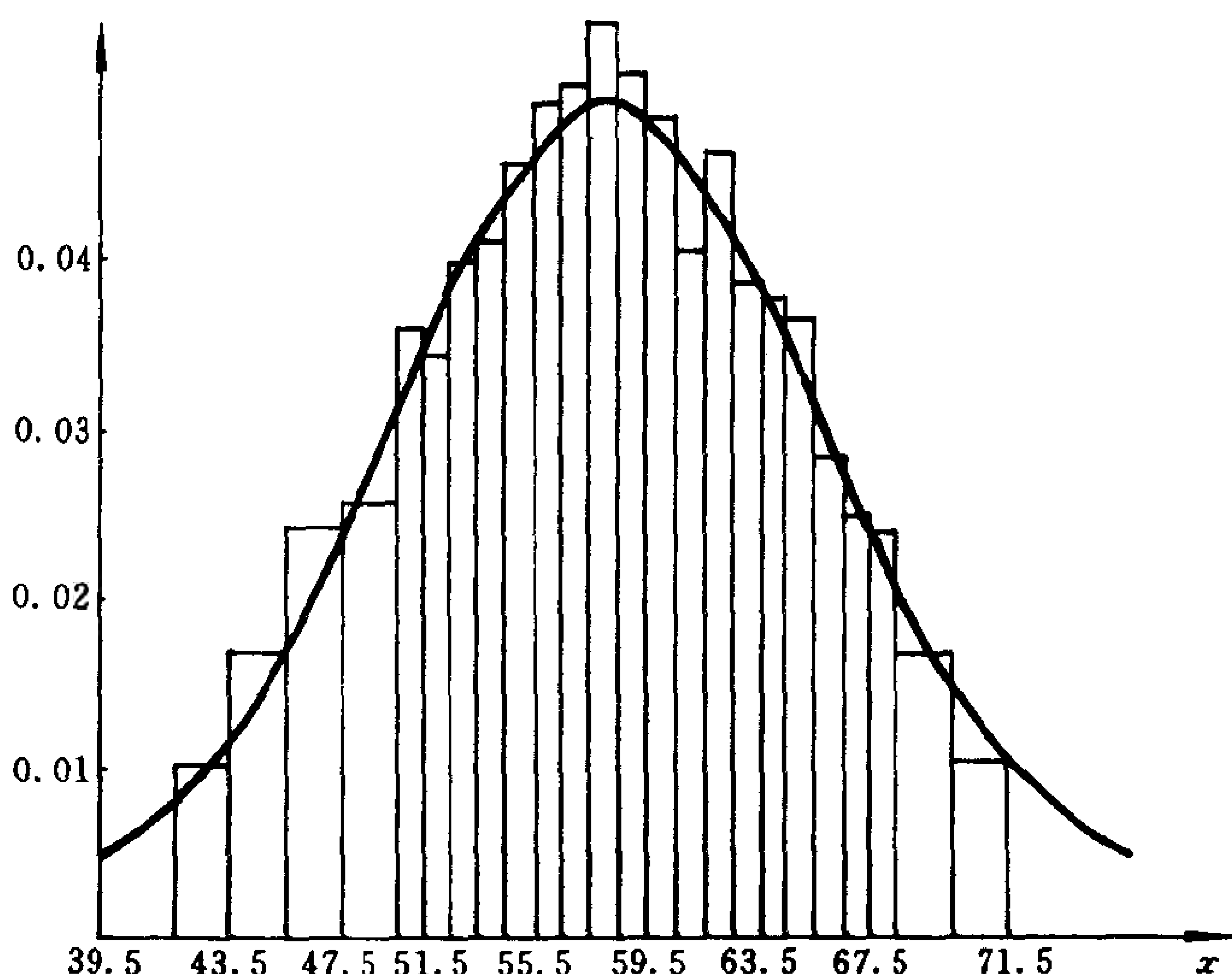


图 21 实测数据频率直方图与正态密度曲线

区间 $[x_i, x_{i+1})$	频数 n_i	频率 $f_i=\frac{n_i}{n}$	$\Phi\left(\frac{x_{i+1}-a}{\sigma}\right)-\Phi\left(\frac{x_i-a}{\sigma}\right)$
$(-\infty, 41.5)$	125	0.03285	0.03005
$[41.5, 43.5)$	72	0.01892	0.02150
$[43.5, 45.5)$	124	0.03259	0.02921
$[45.5, 47.5)$	145	0.03811	0.04431
$[47.5, 49.5)$	193	0.05072	0.05630
$[49.5, 50.5)$	137	0.0360	0.0334
$[50.5, 51.5)$	131	0.0344	0.0398
$[51.5, 52.5)$	154	0.0405	0.0400
$[52.5, 53.5)$	156	0.0410	0.0426
$[53.5, 54.5)$	174	0.0457	0.0457
$[54.5, 55.5)$	186	0.0489	0.0472
$[55.5, 56.5)$	191	0.0502	0.0484
$[56.5, 57.5)$	206	0.0541	0.0486
$[57.5, 58.5)$	193	0.0507	0.0482
$[58.5, 59.5)$	185	0.0486	0.0472
$[59.5, 60.5)$	153	0.0402	0.0454
$[60.5, 61.5)$	176	0.0463	0.0430
$[61.5, 62.5)$	147	0.0386	0.0402
$[62.5, 63.5)$	144	0.0378	0.0370
$[63.5, 64.5)$	140	0.0368	0.0331
$[64.5, 65.5)$	109	0.0286	0.0299
$[65.5, 66.5)$	111	0.0292	0.0282
$[66.5, 67.5)$	93	0.02444	0.02247
$[67.5, 69.5)$	127	0.03338	0.03552
$[69.5, 71.5)$	81	0.02129	0.02547
$[71.5, \infty)$	152	0.03995	0.03754
总　和	$n=3805$		

我们通过下面例子来说明正态分布的计算.

[例 3] 从南郊某地乘车前往北区火车站搭火车有两条路线可走,第一条路线穿过市区,路程较短,但交通拥挤,所需时间(单位为分)服从正态分布 $N(50,100)$,第二条路线沿环城公路走,路程较长,但意外阻塞较少,所需时间服从正态分布 $N(60,16)$,(1)假如有 70 分钟可用,问应走哪一条路线?(2)若只有 65 分钟可用,又应走哪一条路线?

[解] 显然应走在允许的时间内有较大概率及时赶到火车站的路线,若以 τ 记行车时间,则有关概率如下:

(1)有 70 分钟可用时走第一条路线及时赶到的概率为

$$P\{\tau\leqslant 70\}=\Phi\left(\frac{70-50}{10}\right)=\Phi(2)=0.9772$$

走第二条路线及时赶到的概率为

$$P\{\tau\leqslant 70\}=\Phi\left(\frac{70-60}{4}\right)=\Phi(2.5)=0.9938$$

因此在这种场合,应走第二条路线.

(2)只有 65 分钟可用时走第一条路线及时赶到的概率为

$$P\{\tau\leqslant 65\}=\Phi\left(\frac{65-50}{10}\right)=\Phi(1.5)=0.9332$$

走第二条路线及时赶到的概率为

$$P\{\tau\leqslant 65\}=\Phi\left(\frac{65-60}{4}\right)=\Phi(1.25)=0.8944$$

因此这种场合应走第一条路线.

[指数分布] 分布密度函数为

$$p(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}\tag{3.1.41}$$

分布函数为

$$F(x)=\begin{cases}1-e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}\tag{3.1.42}$$

这里 $\lambda>0$,是常数,这分布称为**指数分布**.

指数分布有重要应用，常用它来作为各种“寿命”分布的近似，例如无线电元件的寿命，动物的寿命，电话问题中的通话时间，随机服务系统中的服务时间等都常假定服从指数分布.

指数分布的重要性还表现在它具有类似于几何分布的“无记忆性”. 设随机变量 ξ 服从指数分布(3.1.42)，则对于任意的 $s>0$，$t>0$，

$$P\{\xi\geqslant s+t\mid\xi\geqslant s\}=\frac{P\{\xi\geqslant s+t\}}{P\{\xi\geqslant s\}}=\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}}=e^{-\lambda t}$$

因此

$$P\{\xi\geqslant s+t\mid\xi\geqslant s\}=P\{\xi\geqslant t\} \tag{3.1.43}$$

假如把 ξ 解释为寿命，则(3.1.43)表明，如果已知某人的年龄为 s，则再活 t 年的概率与年龄 s 无关，所以有时又风趣地称指数分布是“永远年青”的.

下列事实也是正确的：指数分布是唯一具有性质(3.1.43)的连续型分布.

我们来证明这个论断.

设 ξ 是非负的，其分布函数为 $F(x)$，记

$$G(x)=P\{\xi\geqslant x\}$$

则由(3.1.43)可以得到

$$G(s+t)=G(s)G(t)$$

对一切 $s\geqslant0,t\geqslant0$ 成立. 因为 $G(x)$ 关于 x 单调，所以由引理 2.4.1 知

$$G(x)=a^{x},\quad x\geqslant0$$

由于 $G(x)$ 是概率，故 $0<a<1$，可以写成 $a=e^{-\lambda}$，其中 $\lambda>0$. 因此

$$F(x)=1-G(x)=1-e^{-\lambda x},\quad x\geqslant0$$

从而证明了结论.

应当指出，指数分布与泊松过程有密切关系，若以 $\xi(t)$ 记参数为 λt 的泊松过程，以 τ_1 记它第一个跳跃发生的时刻，则

$$P\{\tau_1\geqslant t\}=P\{\xi(t)=0\}$$

因此

$$P\{\tau_1 \geqslant t\} = e^{-\lambda t}$$

这说明 τ_1 服从指数分布(3.1.42). 下面把这个结果推广到更为一般的场合.

[埃尔兰分布] 若 $\xi(t)$ 是参数为 λt 的泊松过程, 以 W_r 记它的第 r 个跳跃发生的时刻. 事件 $\{W_r < t\}$ 发生表明第 r 个跳跃出现在时刻 t 之前, 因此事件 $\{\xi(t) \geqslant r\}$ 发生, 即 $\{W_r < t\} \subset \{\xi(t) \geqslant r\}$; 反之, 若事件 $\{\xi(t) \geqslant r\}$ 发生, 即在时刻 t 时 $\xi(t)$ 之值不小于 r, 这时第 r 个跳跃已经出现过, 因此事件 $\{W_r < t\}$ 发生, 即有 $\{\xi(t) \geqslant r\} \subset \{W_r < t\}$. 综上所述可知

$$\{W_r < t\} = \{\xi(t) \geqslant r\} \tag{3.1.44}$$

以 $F(x)$ 记 W_r 的分布函数, 则

$$\begin{aligned} F(t) &= P\{W_r < t\} = P\{\xi(t) \geqslant r\} \\ &= \sum_{k=r}^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} = 1 - \sum_{k=0}^{r-1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \end{aligned}$$

因此

$$\begin{aligned} p(t) &= F'(t) \\ &= -\left[\sum_{k=0}^{r-1} \frac{(\lambda t)^k e^{-\lambda t}(-\lambda)}{k!} + \sum_{k=0}^{r-1} \frac{k(\lambda t)^{k-1} \cdot \lambda e^{-\lambda t}}{k!}\right] \\ &= \lambda e^{-\lambda t} \sum_{k=0}^{r-1} \frac{(\lambda t)^k}{k!} - \lambda e^{-\lambda t} \sum_{k=1}^{r-1} \frac{(\lambda t)^{k-1}}{(k-1)!} \\ &= \frac{\lambda(\lambda t)^{r-1}}{(r-1)!} e^{-\lambda t} = \frac{\lambda^r}{\Gamma(r)} t^{r-1} e^{-\lambda t} \end{aligned}$$

因对于任意的 $r > 0, \lambda > 0$,

$$\int_0^{\infty} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx = 1 \tag{3.1.45}$$

所以, 对任意的正整数 r 及实数 $\lambda > 0$,

$$p(x) = \frac{\lambda^r}{(r-1)!} x^{r-1} e^{-\lambda x}, \qquad x \geqslant 0 \tag{3.1.46}$$

是一个密度函数, 称为**埃尔兰分布**, 它是丹麦科学家埃尔兰

(Erlang)在研究电话问题时引进的，这些研究开创了排队论这一学科.

前面的推导说明，泊松过程中第 r 个跳跃发生的时刻 W_r 服从埃尔兰分布.

当 $r=1$ 时，埃尔兰分布化作指数分布. 另外，若记

$$\tau_1=W_1 \tag{3.1.47}$$
$$\tau_r=W_r-W_{r-1}, \quad r=2,3,\cdots$$

则 τ_r 表示泊松过程的第 r 个来到间隔，用它们可以给出来到时刻 W_r 的如下表达式

$$W_r=\tau_1+\tau_2+\cdots+\tau_r \tag{3.1.48}$$

可以证明，$\tau_1,\tau_2,\cdots,\tau_r$ 均服从参数为 λ 的指数分布，且相互独立.

这个性质与巴斯卡分布类似.

埃尔兰分布也可略加推广，当 $r>0$ 为任意实数时，由(3.1.45)可知能定义如下分布.

［Γ 分布］ 称密度函数为

$$f(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}\mathrm{e}^{-\lambda x}, & x>0\\ 0, & x\leqslant 0\end{cases} \tag{3.1.49}$$

的分布为 **Γ 分布**，其中 $\lambda>0, r>0$ 为参数. 简记作 $G(\lambda,r)$(图 22). 这里，λ 称为尺度参数，r 称为形状参数.

当然，Γ 分布包含埃尔兰分布作为特例，此外，它在概率论和数理统计中还有许多应用，是重要分布之一.

［伯努利过程与泊松过程］ 若每隔 Δt 进行一次试验，则伯努利试验也可以看作一个随时间而变化的过程.

在伯努利试验中，到时刻 $n\Delta t$ 为止，共进行 n 次试验，这时成功次数服从二项分布. 而在泊松过程中，到时刻 t 的来到数则服从泊松分布.

为等待第一次成功，伯努利试验中的等待时间服从几何分布；而泊松过程中则服从指数分布.

为等待第 r 次成功，伯努利试验中的等待时间服从巴斯卡分布；而泊松过程中则服从埃尔兰分布.

正如二项分布的泊松逼近，上述结果可以严格化.不过在这里列出这个对比，只是想让读者有一种系统地记忆几种常见离散型分布和连续型分布的方法.

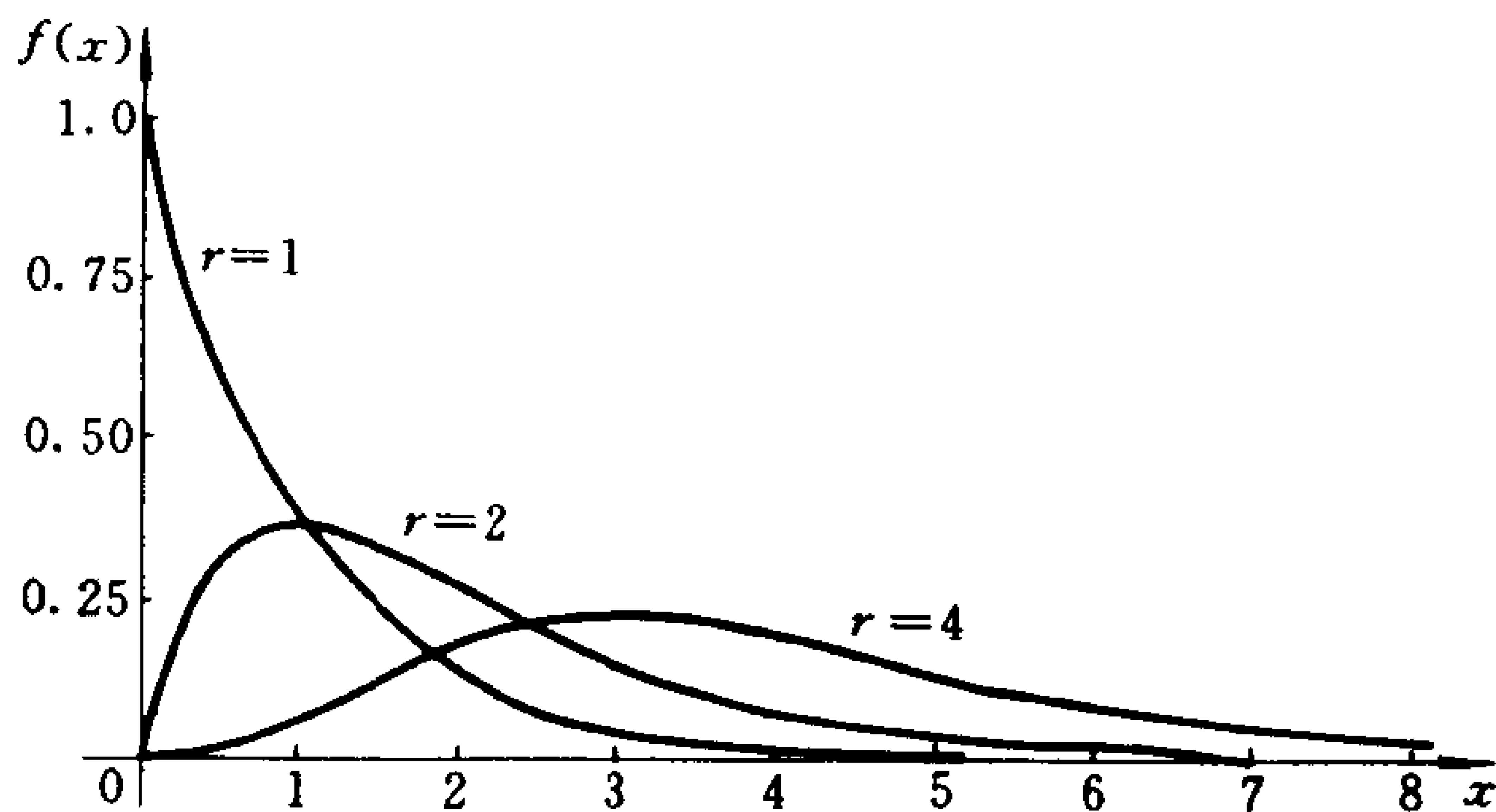


图 22 Γ分布密度函数

*五、关于分布函数的一些结论

我们已经证明了分布函数是单调函数，利用实变数函数论中关于单调函数的一般结果①，不难推出分布函数具有如下性质：

(1) 分布函数至多只有可列个不连续点；

(2) 对分布函数 $F(x)$ 有勒贝格分解

$$F(x)=c_1F_1(x)+c_2F_2(x)+c_3F_3(x) \qquad (3.1.50)$$

其中 $F_1(x)$ 是跳跃函数，$F_2(x)$ 是绝对连续函数，$F_3(x)$ 是所谓奇异函数，它们都是分布函数；而 $0\leqslant c_i\leqslant 1, i=1,2,3$，且 $c_1+c_2+c_3=1$.

在我们讨论过的分布函数中，离散型分布函数是跳跃函数，相当于在(3.1.50)中取 $c_1=1, c_2=c_3=0$ 的场合；而连续型分布函数

① 参看复旦大学数学系. 实变数函数论与泛函分析概要（第二版）. 上海科技出版社. 1963. 第四章

是绝对连续函数，相当于在(3.1.50)中取 $c_2=1,c_1=c_3=0$ 的场合；自然会想到，也可以取 $c_3=1,c_1=c_2=0$，得到另一类分布函数. 这个结论是正确的，理论上确实存在着另一类分布函数——奇异型分布函数，它是连续函数，但却不能表为不定积分，因此它没有密度函数. 不过到目前为止，常用的分布都是离散型或连续型的，因此我们不准备对奇异型分布多加讨论. 以后证明结果或是对离散型进行，或是对连续型进行，或者对一般分布进行.

§2. 随机向量，随机变量的独立性

一、随机向量及其分布

在有些随机现象中，每次试验的结果不能只用一个数来描述，而要同时用几个数来描述，例如对于钢的成份，需要同时指出它的含碳量、含硫量、含磷量等等. 这样，对应于每个样本点 ω，试验的结果将是一个向量 $(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega))$，这个向量取值于 n 维欧几里得空间 R^n.

定义 3.2.1 若随机变量 $\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega)$ 定义在同一概率空间 $(\Omega,\mathscr{F},P)$ 上，则称

$$\xi(\omega)=(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega)) \tag{3.2.1}$$

构成一个 **n 维随机向量**，亦称 **n 维随机变量**. 显然，一维随机向量即为随机变量.

固然可以对随机向量的一个个分量分别研究，但是我们马上就会看到，把它们作为一个向量，则不但能研究各个分量的性质，而且还可以考察它们之间的联系，对许多问题来说，这是十分必要的.

对于任意的 n 个实数 $x_1,x_2,\cdots,x_n$，

$$\begin{aligned}&\{\xi_1(\omega)<x_1,\xi_2(\omega)<x_2,\cdots,\xi_n(\omega)<x_n\}\\&=\bigcap_{i=1}^{n}\{\xi_i(\omega)<x_i\}\in\mathscr{F}\end{aligned} \tag{3.2.2}$$

亦即对于 R^n 中的 n 维矩形 $C_n=\prod_{i=1}^{n}(-\infty,x_i)$，有

$$\{\xi(\omega)\in C_n\}\in\mathscr{F}$$

利用测度论的方法还可证明，若 B_n 为 R^n 上任一博雷尔点集，也有

$$\{\xi(\omega)\in B_n\}\in\mathscr{F} \tag{3.2.3}$$

以后我们将要用到这个结论(图 23).

类似于一维的场合，我们引进如下定义.

定义 3.2.2 称 n 元函数

$$F(x_1,x_2,\cdots,x_n)=P(\xi_1(\omega)<x_1,\xi_2(\omega)<x_2,\cdots,\xi_n(\omega)<x_n\} \tag{3.2.4}$$

为随机向量 $\xi(\omega)=(\xi_1(\omega),\xi_2(\omega),\cdots,\xi_n(\omega))$ 的(**联合**)**分布函数**.

给定了联合分布函数后，可以计算事件 $\{a_1\leqslant\xi_1<b_1,a_2\leqslant\xi_2<b_2,\cdots,a_n\leqslant\xi_n<b_n\}$ 的概率，例如当 $n=2$ 时，

$$\begin{aligned}&P\{a_1\leqslant\xi_1<b_1,a_2\leqslant\xi_2<b_2\}\\&=F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2)\end{aligned} \tag{3.2.5}$$

这个结果容易从图 24 看出.

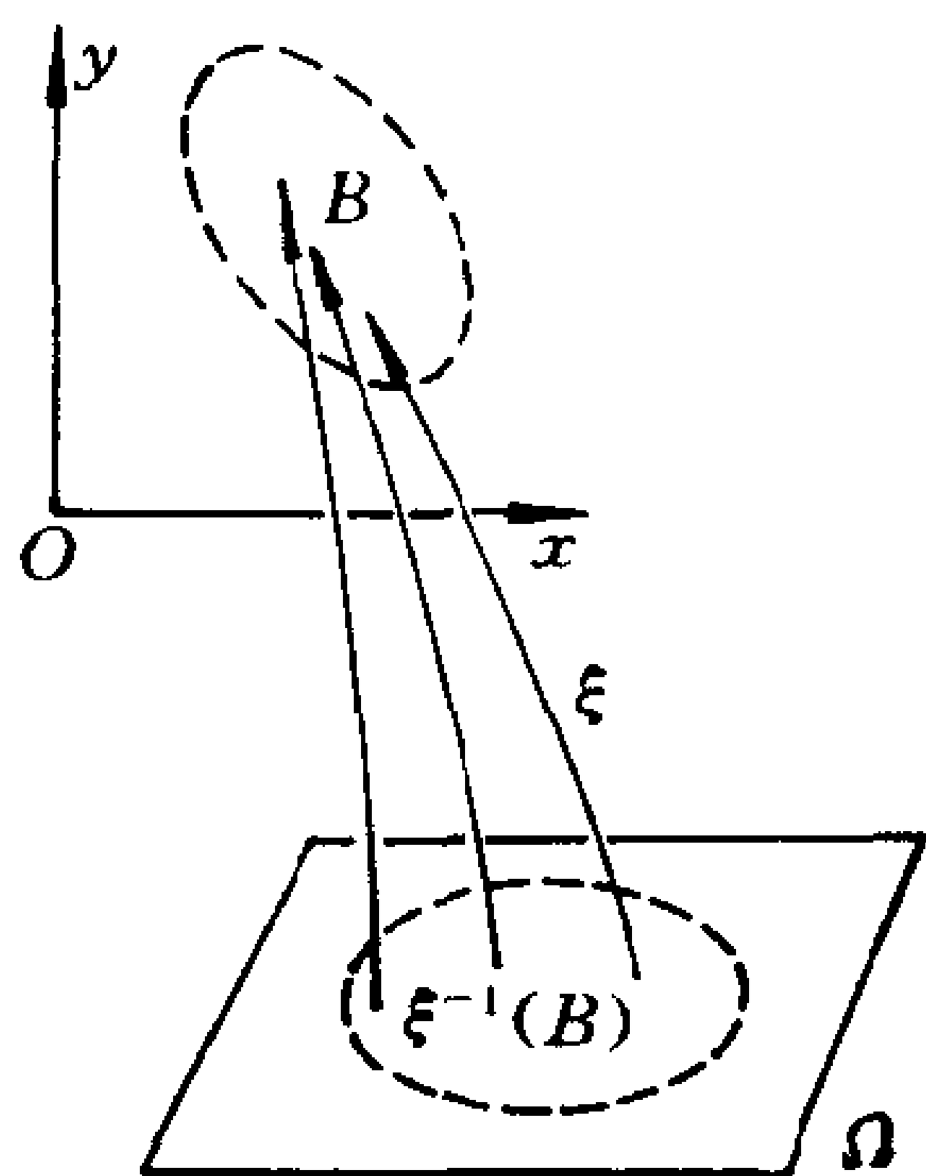


图 23　二维随机向量

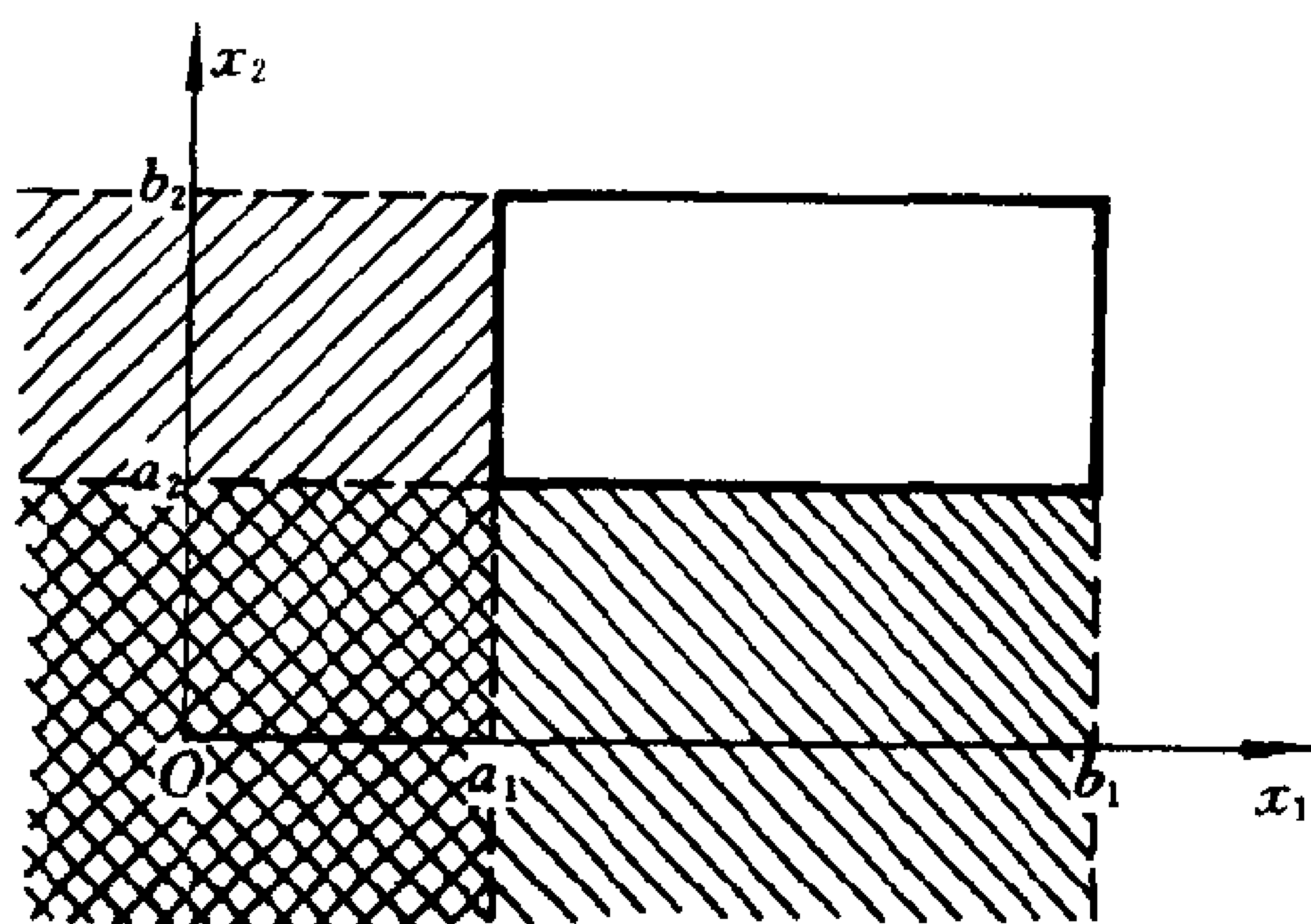


图 24　二维概率计算

类似于一元的场合，可以证明多元分布函数的一些性质.

(i) 单调性：关于每个变元是单调不减函数；

(ii) $F(x_1,x_2,\cdots,-\infty,\cdots,x_n)=0$,

$F(+\infty,+\infty,\cdots,+\infty)=1$;

(iii) 关于每个变元左连续.

在二元场合，还应该有：

(iv) 对任意 $a_1<b_1,a_2<b_2$，都有

$$F(b_1,b_2)-F(a_1,b_2)-F(b_1,a_2)+F(a_1,a_2)\geqslant 0$$

为保证(3.2.5)式中的概率的非负性，性质(iv)是必须的，而且由性质(iv)可以推出单调性，但存在着反例说明，由单调性并不能保证性质(iv)的成立(见习题 12). 这是多元场合与一元场合的不同之处.

可以证明：满足(ii),(iii),(iv)这三条性质的二元函数是某二维随机变量的分布函数.

类似的结论对 n 元场合也成立.

随机向量也有不同类型，最常见的也是离散型与连续型两类.

在离散型场合，概率分布集中在有限或可列个点上. 重要的多元离散型分布有多项分布与多元超几何分布，它们分别是二项分布与超几何分布往多元场合的推广.

[多项分布] 在试验中，若每次试验的可能结果为 $A_1,A_2,\cdots,A_r$，而 $P(A_i)=p_i,i=1,2,\cdots,r$，且 $p_1+p_2+\cdots+p_r=1$，重复这种试验 n 次，并假定这些试验是相互独立的，若以 $\xi_1,\xi_2,\cdots,\xi_r$ 分别记 $A_1,A_2,\cdots,A_r$ 出现的次数，则

$$P\{\xi_1=k_1,\xi_2=k_2,\cdots,\xi_r=k_r\}=\frac{n!}{k_1!\,k_2!\cdots k_r!}p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r} \tag{3.2.6}$$

这里整数 $k_i\geqslant 0$，且仅当 $k_1+k_2+\cdots+k_r=n$ 时上式才成立，否则为 0. 多项分布在第二章 § 3 中已出现过.

[多元超几何分布] 袋中装 i 号球 N_i 只，$i=1,2,\cdots,r$，$N_1+N_2+\cdots+N_r=N$，从中随机摸出 n 只，若以 $\xi_1,\xi_2,\cdots,\xi_r$ 分别

记 1,2,…,r 号球的出现数,则

$$P\{\xi_1=n_1,\xi_2=n_2,\cdots,\xi_r=n_r\}=\frac{\binom{N_1}{n_1}\binom{N_2}{n_2}\cdots\binom{N_r}{n_r}}{\binom{N}{n}}\tag{3.2.7}$$

这里整数 $n_i\geqslant 0$,且仅当 $n_1+n_2+\cdots+n_r=n$ 时上式才成立,否则为 0. 古典概型一节的例 2 就是这种类型.

以上两个分布在抽样中常用,前者用于有放回场合,后者则用于不放回场合.

在连续型场合,存在着非负函数 $p(x_1,x_2,\cdots,x_n)$,使

$$F(x_1,\cdots,x_n)=\int_{-\infty}^{x_1}\cdots\int_{-\infty}^{x_n}p(y_1,\cdots,y_n)\mathrm{d}y_1\cdots\mathrm{d}y_n\tag{3.2.8}$$

这里的 $p(x_1,\cdots,x_n)$ 称为**密度函数**,满足如下两个条件

$$p(x_1,\cdots,x_n)\geqslant 0\tag{3.2.9}$$

$$\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n=1\tag{3.2.10}$$

随机向量的概念在各个基础学科和工程技术中已有广泛应用. 例如在量子力学中,粒子在某个区域 G 的出现是通过概率来描述的. 若以 ψ 表示它的波函数,则 $|\psi|^2$ 即为密度函数,而

$$\iiint_G |\psi|^2\mathrm{d}x\mathrm{d}y\mathrm{d}z$$

就给出该粒子在区域 G 出现的概率.

均匀分布和 n 元正态分布是比较常见的两种多维连续型分布.

[均匀分布] 若 G 为 $\mathbf{R}^n$ 中有限区域,其测度 $S>0$;则由密度函数

$$p(x_1,\cdots,x_n)=\begin{cases}\dfrac{1}{S}, & (x_1,\cdots,x_n)\in G\\ 0, & (x_1,\cdots,x_n)\overline{\in} G\end{cases}\tag{3.2.11}$$

给出的分布称为**G 上的均匀分布**.

在第一章几何概率一节中，我们已看到均匀分布的各种例子.

[多元正态分布] 若 $\boldsymbol{B}=(b_{ij})$ 是 n 阶正定对称矩阵，以 $\boldsymbol{B}^{-1}=(r_{ij})$ 表示 $\boldsymbol{B}$ 的逆阵；$\det\boldsymbol{B}$ 表示 $\boldsymbol{B}$ 的行列式的值. $\boldsymbol{a}=(a_1,\cdots,a_n)$ 是任意实值行向量，则由密度函数

$$p(x_1,\cdots,x_n)=\frac{1}{(2\pi)^{\frac{n}{2}}(\det\boldsymbol{B})^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\sum_{j,k=1}^{n}r_{jk}(x_j-a_j)(x_k-a_k)\right\} \tag{3.2.12}$$

定义的分布称为**n 元正态分布**，简记为 $N(\boldsymbol{a},\boldsymbol{B})$.

这个密度函数也可以改为如下向量形式：

$$p(x)=\frac{1}{(2\pi)^{\frac{n}{2}}(\det\boldsymbol{B})^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})'\right\} \tag{3.2.13}$$

这里 $(\boldsymbol{x}-\boldsymbol{a})'$ 表示向量 $(\boldsymbol{x}-\boldsymbol{a})$ 的转置.

n 元正态分布是最重要的一种多维分布，它在概率论、数理统计、随机过程论中都占有重要地位，具有许多重要性质. 对于这些性质的叙述和证明，我们将在引进了更有力的工具后再进行. 这里我们将对它的特殊场合——二元正态分布逐步直接导出这些性质.

二、边际分布

为方便起见，讨论将对二维场合进行，多维时这些结论仍然成立.

先讨论离散型分布的场合，在这种场合，有关概念特别容易理解.

考虑二维随机向量 (ξ,η)，设 ξ 取值 x_1，x_2，$\cdots$；η 取值 y_1，y_2，$\cdots$记

$$P\{\xi=x_i,\eta=y_j\}=p(x_i,y_j),\quad i,j=1,2,\cdots \tag{3.2.14}$$

$$P\{\xi=x_i\}=p_1(x_i),\qquad i=1,2,\cdots \tag{3.2.15}$$

$$P\{\eta=y_j\}=p_2(y_j),\qquad j=1,2,\cdots \tag{3.2.16}$$

显然

$$p(x_i,y_j)\geqslant 0,\qquad \sum_{i,j}p(x_i,y_j)=1 \tag{3.2.17}$$

此外，对固定的 i，

$$\sum_j p(x_i,y_j)=P\{\xi=x_i\}=p_1(x_i) \tag{3.2.18}$$

而对固定的 j，

$$\sum_i p(x_i,y_j)=P\{\eta=y_j\}=p_2(y_j) \tag{3.2.19}$$

换句话说，由联合概率分布，对于固定的 i 关于 j 求和得到 ξ 的概率分布；而对于固定的 j 关于 i 求和得到 η 的概率分布.

这里 $p_1(x_i)$ 与 $p_2(y_j)$ 称为 $p(x_i,y_j)$ 的**边际分布**. 这个名称的含义通过下列两例将看得很清楚.

[例 1] 袋中装有 2 只白球及 3 只黑球. 现进行有放回的摸球，定义下列随机变量：

$$\xi=\begin{cases}1, \text{第 1 次摸出白球}\\ 0, \text{第 1 次摸出黑球}\end{cases}\qquad \eta=\begin{cases}1, \text{第 2 次摸出白球}\\ 0, \text{第 2 次摸出黑球}\end{cases}$$

则 (ξ,η) 的联合概率分布与边际分布由下表给出：

η \ ξ	0	1	$p_2(y_j)$
0	$\frac{3}{5}\cdot\frac{3}{5}$	$\frac{2}{5}\cdot\frac{3}{5}$	$\frac{3}{5}$
1	$\frac{3}{5}\cdot\frac{2}{5}$	$\frac{2}{5}\cdot\frac{2}{5}$	$\frac{2}{5}$
$p_1(x_i)$	$\frac{3}{5}$	$\frac{2}{5}$	

[例 2] 前例中若采用不放回摸球，则 (ξ,η) 的联合概率分布及边际分布由下表给出：

η \ ξ	0	1	$p_2(y_j)$
0	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{3}{4}$	$\frac{3}{5}$
1	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{1}{4}$	$\frac{2}{5}$
$p_1(x_i)$	$\frac{3}{5}$	$\frac{2}{5}$	

在上面两个表中，中间部分是 (ξ,η) 的联合概率分布，而边沿部分是 ξ 及 η 的概率分布，它们由联合分布经同一行或同一列的相加而得出来，这种表称为列联表. 在列联表中，ξ 与 η 的概率分布处于表的边沿部位，因此称为边际分布.

让我们再注意一个重要事实，两例中 ξ 及 η 的（边际）分布是相同的，但是它们的联合分布却完全不同. 这里可以看出联合分布不能由边际分布唯一确定，也就是说二维随机向量的性质并不能由它两个分量的个别性质来确定，这时还必须考虑它们之间的联系，这也说明了研究多维随机向量的作用.

一般地，若 (ξ,η) 是二维随机向量，其分布函数为 $F(x,y)$，我们能由 $F(x,y)$ 得出 ξ 或 η 的分布函数. 事实上，

$$F_1(x) = P\{\xi < x\} = P\{\xi < x,\eta < +\infty\} = F(x,+\infty) \tag{3.2.20}$$

同理

$$F_2(y) = P\{\eta < y\} = F(+\infty,y)$$

$F_1(x)$ 及 $F_2(y)$ 称为 $F(x,y)$ 的 **边际分布函数**.

若 $F(x,y)$ 是连续型分布函数，有密度函数 $p(x,y)$，那么

$$F_1(x) = \int_{-\infty}^{x}\int_{-\infty}^{\infty} p(u,y)\mathrm{d}u\mathrm{d}y$$

因此 $F_1(x)$ 是连续型分布函数，其密度函数为

$$p_1(x) = \int_{-\infty}^{\infty} p(x,y)\mathrm{d}y \tag{3.2.21}$$

同理 $F_2(x)$是连续型分布函数，其密度函数为

$$p_2(y) = \int_{-\infty}^{\infty} p(x,y)\mathrm{d}x$$

$p_1(x)$ 及 $p_2(y)$ 称为 $p(x,y)$ 的**边际分布密度函数**.

［二元正态分布］ 函数

$$p(x,y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)} \cdot \left[\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2}\right]\right\} \tag{3.2.22}$$

这里 $a_1, a_2, \sigma_1, \sigma_2, r$ 为常数，$\sigma_1 > 0, \sigma_2 > 0, |r| < 1$，称为**二元正态分布密度函数**（图 25）.

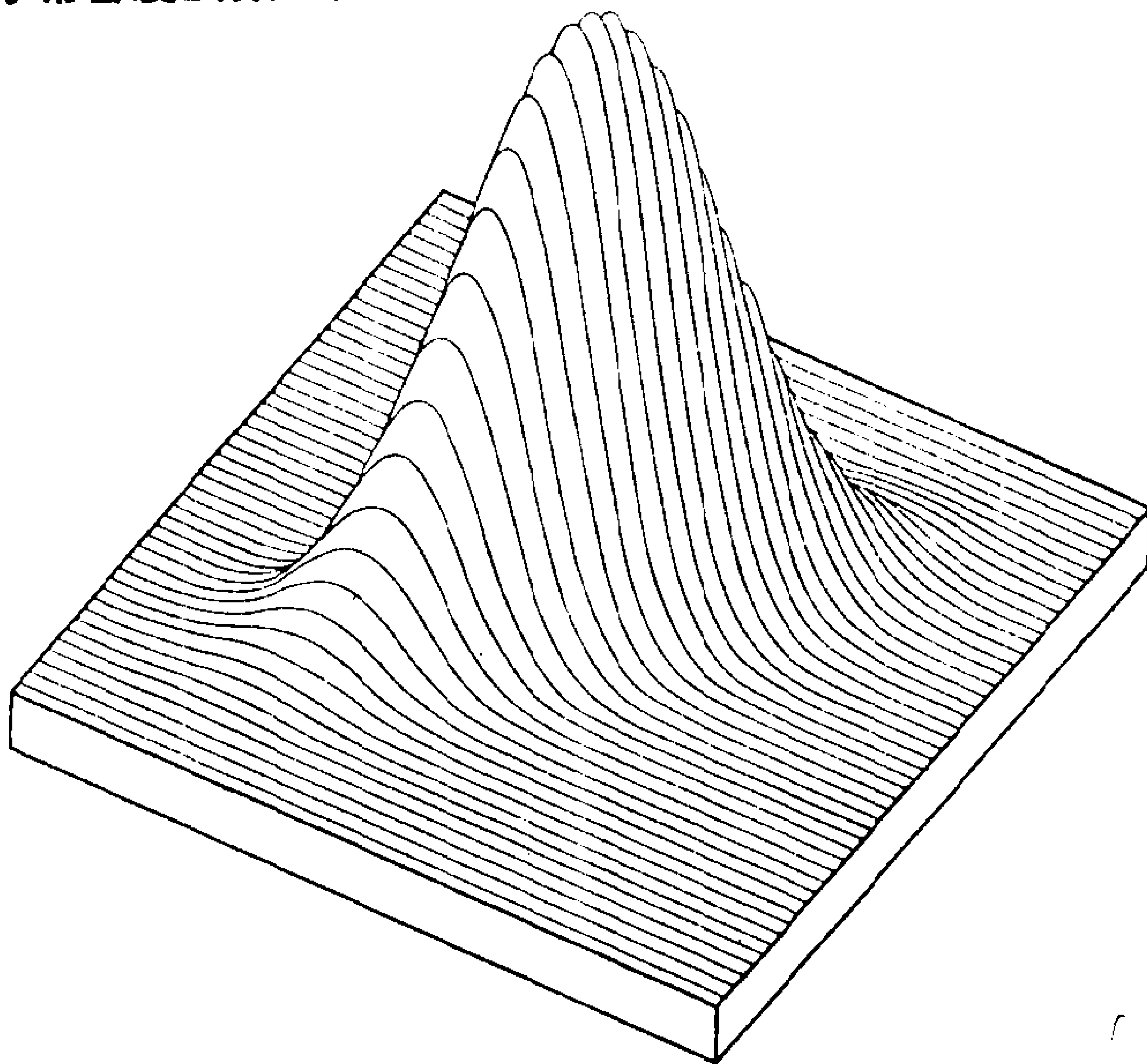


图 25 二维正态密度曲面

显然这是n元正态分布当$n=2$时的特殊情况，相应的

$$\boldsymbol{B}=\begin{pmatrix}\sigma_1^2 & r\sigma_1\sigma_2\\ r\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix},\quad \boldsymbol{a}_1=a_1,\boldsymbol{a}_2=a_2$$

我们来求其边际分布密度函数，令$\dfrac{x-a_1}{\sigma_1}=u,\dfrac{y-a_2}{\sigma_2}=v$，由(3.2.21)知

$$\begin{aligned}p_1(x)&=\int_{-\infty}^{\infty}p(x,y)\mathrm{d}y\\&=\frac{1}{2\pi\sigma_1\sqrt{1-r^2}}\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2(1-r^2)}\right.\\&\quad\cdot\left.[u^2-2ruv+v^2]\right\}\mathrm{d}v\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-u^2/2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\\&\quad\cdot\exp\left(-\frac{r^2u^2-2ruv+v^2}{2(1-r^2)}\right)\mathrm{d}v\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-u^2/2}\cdot\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi(1-r^2)}}\mathrm{e}^{-(v-ru)^2/[2(1-r^2)]}\mathrm{d}v\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-u^2/2}\\&=\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-a_1)^2}{2\sigma_1^2}}\end{aligned}\tag{3.2.23}$$

即$p_1(x)$是$N(a_1,\sigma_1^2)$的密度函数. 同理

$$p_2(y)=\frac{1}{\sqrt{2\pi}\sigma_2}\mathrm{e}^{-\frac{(y-a_2)^2}{2\sigma_2^2}}\tag{3.2.24}$$

因此二元正态分布的边际分布仍为正态分布，这是一个重要的结论.

由(3.2.22)定义的$p(x,y)$显然非负，又由于

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}p(x,y)\mathrm{d}x\mathrm{d}y=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_1}\mathrm{e}^{-\frac{(x-a_1)^2}{2\sigma_1^2}}\mathrm{d}x=1$$

因此我们已顺带验证了$p(x,y)$确实是一个密度函数.

对于 n 维场合可以类似讨论其边际分布，值得提醒的是，对 n 维分布而言，存在着 $n-1$ 维，$n-2$ 维，……2 维，1 维的边际分布.

三、条件分布

对于多个随机事件可以讨论它们的条件概率，同样地，对于多个随机变量也可以讨论它们的条件分布，并由此得出重要结果.

仍对二维的场合进行讨论. 也还是先从离散型开始，这时并无多大困难.

若已知 $\xi = x_i(p_1(x_i) > 0)$，则事件 $\{\eta = y_j\}$ 的条件概率为

$$P\{\eta = y_j | \xi = x_i\} = \frac{P\{\xi = x_i, \eta = y_j\}}{P\{\xi = x_i\}} = \frac{p(x_i, y_j)}{p_1(x_i)} \tag{3.2.25}$$

这式子定义了随机变量 η 关于随机变量 ξ 的**条件分布**. 在一般情况下，它不同于 $p_2(y_j)$，这表示从 ξ 的取值可以得出关于 η 的某些信息，反之亦然.

对于一般随机向量 (ξ, η)，我们也想定义条件分布函数 $P\{\eta < y | \xi = x\}$，但是由于会出现 $P\{\xi = x\} = 0$，因此我们不能像(3.2.25)那样简单地定义.

自然会想到可以用

$$\begin{aligned} P\{\eta < y | \xi = x\} &= \lim_{\Delta x \to 0} P\{\eta < y | x \leqslant \xi < x + \Delta x\} \\ &= \lim_{\Delta x \to 0} \frac{P\{x \leqslant \xi < x + \Delta x, \eta < y\}}{P\{x \leqslant \xi < x + \Delta x\}} \\ &= \lim_{\Delta x \to 0} \frac{F(x + \Delta x, y) - F(x, y)}{F(x + \Delta x, \infty) - F(x, \infty)} \end{aligned} \tag{3.2.26}$$

来定义.

特别是对于有连续密度函数的场合，这定义导出

$$P\{\eta < y | \xi = x\} = \lim_{\Delta x \to 0} \frac{\int_x^{x+\Delta x} \int_{-\infty}^{y} p(u, v) \mathrm{d}u \mathrm{d}v}{\int_x^{x+\Delta x} \int_{-\infty}^{\infty} p(u, v) \mathrm{d}u \mathrm{d}v}$$

若把上式的分子分母分别除以 Δx，再令 $\Delta x \to 0$ 取极限，则当 $p_1(x) \neq 0$ 时，

$$P\{\eta < y \mid \xi = x\} = \frac{\int_{-\infty}^{y} p(x,v)\mathrm{d}v}{p_1(x)} = \int_{-\infty}^{y} \frac{p(x,v)}{p_1(x)}\mathrm{d}v$$

因此在给定 $\xi = x$ 的条件下，η 的分布密度函数为

$$p(y \mid x) = \frac{p(x,y)}{p_1(x)} \tag{3.2.27}$$

同理可得在给定 $\eta = y$ 的条件下，ξ 的分布密度函数为

$$p(x \mid y) = \frac{p(x,y)}{p_2(y)} \tag{3.2.28}$$

这里当然也要求 $p_2(y) \neq 0$.

［例 3］ 对由（3.2.22）定义的二元正态分布求条件密度函数 $p(y \mid x)$，由（3.2.27）

$$\begin{aligned}
p(y \mid x) = \frac{p(x,y)}{p_1(x)} &= \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}} \exp\Big(-\frac{1}{2(1-r^2)} \\
&\quad \cdot \Big[\frac{(x-a_1)^2}{\sigma_1^2} - \frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} \Big] \\
&\quad + \frac{(x-a_1)^2}{2\sigma_1^2} \Big) \\
&= \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}} \exp\Big\{ -\frac{1}{2(1-r^2)} \\
&\quad \cdot \Big[\frac{(x-a_1)^2 r^2}{\sigma_1^2} - \frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2} + \frac{(y-a_2)^2}{\sigma_2^2} \Big] \Big\} \\
&= \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}} \\
&\quad \cdot \exp\Big\{ -\frac{1}{2(1-r^2)} \Big(\frac{y-a_2}{\sigma_2} - r\frac{x-a_1}{\sigma_1} \Big)^2 \Big\} \\
&= \frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-r^2}} \exp\Big\{ -\frac{1}{2\sigma_2^2(1-r^2)} \\
&\quad \cdot \Big[y - \Big(a_2 + r\frac{\sigma_2}{\sigma_1}(x-a_1) \Big) \Big]^2 \Big\}
\end{aligned} \tag{3.2.29}$$

从这里我们看到，二元正态分布的条件分布仍然是正态分布

$$N\left(a_2 + r\frac{\sigma_2}{\sigma_1}(x - a_1), \sigma_2^2(1 - r^2)\right) \qquad (3.2.30)$$

特别指出，这里 $m = a_2 + r\frac{\sigma_2}{\sigma_1}(x - a_1)$ 是 x 的线性函数，这个结论在一些统计问题中很重要.

四、随机变量的独立性

在上章中，我们看到随机事件的独立性起着很大的作用．下面研究随机变量的独立性．引入如下定义.

定义 3.2.3 设 $\xi_1, \cdots, \xi_n$ 为 n 个随机变量，若对于任意的 $x_1, \cdots, x_n$ 成立

$$P\{\xi_1 < x_1, \cdots, \xi_n < x_n\} = P\{\xi_1 < x_1\} \cdots P\{\xi_n < x_n\} \qquad (3.2.31)$$

则称 $\xi_1, \cdots, \xi_n$ 是**相互独立的**.

若 ξ_i 的分布函数为 $F_i(x), i = 1, 2, \cdots, n$，它们的联合分布函数为 $F(x_1, \cdots, x_n)$，则(3.2.31)等价于对一切 $x_1, \cdots, x_n$ 成立

$$F(x_1, \cdots, x_n) = F_1(x_1) \cdots F_n(x_n) \qquad (3.2.32)$$

在这种场合，由每个随机变量的（边际）分布函数可以唯一地确定联合分布函数．而且由(3.2.26)可以看到，这时条件分布化为无条件分布

$$P\{\eta < y | \xi = x\} = P\{\eta < y\} \qquad (3.2.33)$$

即由 ξ 的取值不能得出任何关于 η 的信息.

对于离散型随机变量，(3.2.31)等价于对任何一组可能取的值 $(x_1, \cdots, x_n)$ 成立

$$P\{\xi_1 = x_1, \cdots, \xi_n = x_n\} = P\{\xi_1 = x_1\} \cdots P\{\xi_n = x_n\} \qquad (3.2.34)$$

例 1 中有放回摸球时的 ξ 与 η 是相互独立的，这时联合分布取乘法表的形式；例 2 中的 ξ 与 η 不是独立的．这些例子也说明，

如果两个随机变量有相同的分布，则它们可以是独立的，也可以不是独立的.

对于连续型随机变量，条件(3.2.31)的等价形式是对 $x_1,\cdots,x_n$ 几乎处处成立

$$p(x_1,\cdots,x_n)=p_1(x_1)\cdots p_n(x_n) \qquad (3.2.35)$$

这里 $p(x_1,\cdots,x_n)$ 是联合分布密度函数，而 $p_i(x_i),i=1,2,\cdots,n$ 是各随机变量的密度函数.

[例 4] 对由(3.2.22)定义的二元正态分布，有

$$p_1(x)p_2(y)=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{(x-a_1)^2}{2\sigma_1^2}-\frac{(y-a_2)^2}{2\sigma_2^2}\right\}$$

与(3.2.22)比较可知，使关系式(3.2.35)成立的充要条件是

$$r=0 \qquad (3.2.36)$$

即服从二元正态分布的随机变量独立的充要条件是 $r=0$，这时条件分布(3.2.29)化为

$$p(y|x)=\frac{1}{\sqrt{2\pi}\sigma_2}\ \mathrm{e}^{\frac{-(y-a_2)^2}{2\sigma_2^2}}=p_2(y)$$

这与(3.2.33)是一致的.

上面我们列举了随机变量独立性的各种表达形式，有些是对一般随机变量成立的，有些只对离散型或连续型才成立，一般说来，这些条件比较易于验证. 下面我们介绍另外一个条件，这条件不易验证，但在理论研究中有用.

随机变量 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立的充要条件是对一切一维博雷尔点集 $A_1,A_2,\cdots,A_n$ 成立

$$P\{\xi_1\in A_1,\xi_2\in A_2,\cdots,\xi_n\in A_n\}$$
$$=P\{\xi_1\in A_1\}P\{\xi_2\in A_2\}\cdots P\{\xi_n\in A_n\} \qquad (3.2.37)$$

论断的证明要用到测度论，已超出本课程范围.

当然也可以建立 n 维随机向量 $\boldsymbol{\xi}$ 与 m 维随机向量 $\boldsymbol{\eta}$ 相互独立的概念，这时要求成立

$$P\{\boldsymbol{\xi}\in A,\boldsymbol{\eta}\in B\}=P\{\boldsymbol{\xi}\in A\}P\{\boldsymbol{\eta}\in B\}$$

其中 A，B 分别是任意一个 n 维及 m 维的博雷尔点集.

显然若 ξ 与 η 独立，则 ξ 的子向量与 η 的子向量是独立的.

此外，注意到若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立,则其中的任意 $r(2\leqslant r<n)$ 个随机变量也相互独立. 例如,我们证明 $\xi_1,\xi_2,\cdots,\xi_{n-1}$ 相互独立.

$$\begin{aligned}&P\{\xi_1<x_1,\cdots,\xi_{n-1}<x_{n-1}\}\\ =&P\{\xi_1<x_1,\cdots,\xi_{n-1}<x_{n-1},\xi_n<\infty\}\\ =&P\{\xi_1<x_1\}\cdots P\{\xi_{n-1}<x_{n-1}\}P\{\xi_n<\infty\}\\ =&P\{\xi_1<x_1\}\cdots P\{\xi_{n-1}<x_{n-1}\}\end{aligned}$$

最后，称无穷多个随机变量是相互独立的，如果其中任意有限多个随机变量都是相互独立的.

随机变量的独立性概念是概率论中最基本的概念之一，也是最重要的概念之一，关于独立随机变量的研究构成了概率论的重要课题，我们将在第五章中介绍一些基本结果.

有了独立性的概念，可以介绍一个导出正态分布的条件. 先证明一个分析引理.

引理 3.2.1 若 $f(x)$ 及 $g(y)(x\geqslant 0,y\geqslant 0)$ 是不恒等于 0 的连续函数，而且对一切非负实数 x 及 y 成立

$$f(x)g(y)=h(x+y) \tag{3.2.38}$$

则

$$f(x)=ka^x,\quad x\geqslant 0 \tag{3.2.39}$$

这里 k 及 a 是常数,$a>0$.

[证明] 首先证明 $f(0)\neq 0$. 用反证法,假定 $f(0)=0$,则 $h(y)\equiv 0$,由于 $g(y)$ 不恒为 0,有一个 y_0 使 $g(y_0)\neq 0$. 因此 $f(x)=\dfrac{h(x+y_0)}{g(y_0)}\equiv 0$，这与 $f(x)$ 不恒为 0 的条件矛盾,因此 $f(0)\neq 0$. 同理可证 $g(0)\neq 0$.

其次在 (3.2.38) 中令 $x=z,y=0$,得 $f(z)g(0)=h(z)$,又在(3.2.38) 中令 $x=0,y=z$ 得 $f(0)g(z)=h(z)$，因此

$$f(z)g(0) = f(0)g(z) = h(z)$$

两边除以 $f(0)g(0)$ 并记得到的比值为 $p(z)$，则得

$$\frac{f(z)}{f(0)} = \frac{g(z)}{g(0)} = \frac{h(z)}{f(0)g(0)} = p(z)$$

代入（3.2.38）得

$$p(x)p(y) = p(x+y)$$

利用引理 2.4.1 知

$$p(x) = a^x$$

因此

$$f(x) = f(0)a^x$$

引理得证.

［射击弹落点分布］ 在火器射击问题中，假如过射击目标中心作直角坐标系的 x 轴及 y 轴，每次射击弹落点由于受随机因素影响会偏离目标，它的坐标 (ξ,η) 是一个二维随机变量. 假定满足下列条件：

（1） ξ 与 η 具有连续的密度函数，分别记为 $p_1(x)$ 及 $p_2(y)$；

（2） ξ 与 η 相互独立；

（3） (ξ,η) 的密度函数在 (x,y) 点的值仅与它到原点的距离 $r = \sqrt{x^2+y^2}$ 有关.

则 ξ 与 η 均服从正态分布.

［证明］ 记 (ξ,η) 的密度函数为 $p(x,y)$，由假定（3）知 $p(x,y) = q(\sqrt{x^2+y^2})$，再由假定（1），（2）知

$$p_1(x)p_2(y) = q(\sqrt{x^2+y^2})$$

因为 $p_1(x)$ 及 $p_2(y)$ 是密度函数，当然不能恒等于 0，故可利用引理 3.2.1［把 $p_1(x)$ 看作 $f(x^2)$，$p_2(y)$ 看作 $g(y^2)$，$q(\sqrt{x^2+y^2})$ 看作 $h(x^2+y^2)$］，得到

$$p_1(x) = ka^{x^2}$$

由于 $p_1(x)$ 是 $(-\infty,\infty)$ 上的密度函数，因此必须有 $0<a<1$，记

$a = e^{-1/(2\sigma_1^2)}$，其中 $\sigma_1 > 0$，利用 $\int_{-\infty}^{\infty} p_1(x)dx = 1$，可知 $k = \frac{1}{\sqrt{2\pi}\sigma_1}$，这样一来，

$$p_1(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{x^2}{2\sigma_1^2}}$$

因此 ξ 服从正态分布. 同理可证 η 也服从正态分布.

§3. 随机变量的函数及其分布

一、博雷尔函数与随机变量的函数

在许多问题中需要计算随机变量的函数的分布律，例如在无线电接收中，某时刻收到的信号是一个随机变量 ξ，若我们把这个信号通过平方检波器，则输出的信号为 $\eta = \xi^2$，这时我们需要求 η 的分布律. 又如在统计物理中，已知分子运动速度 ξ 的分布，要求其动能 $\eta = \frac{1}{2}m\xi^2$ 的分布律. 这类问题既普遍而又重要，接下来我们就要讨论它.

这类问题较为一般的提法是：若 ξ 是随机变量，求 $\eta = g(\xi)$ 的分布律.

为了使 η 有分布律可言，当然要求 η 是随机变量，因此对函数 $y = g(x)$ 也必须有一定的要求，为此我们提出如下定义.

定义 3.3.1 设 $y = g(x)$ 是 R^1 到 R^1 上的一个映照，若对于一切 R^1 中的博雷尔点集 B_1 均有

$$\{x : g(x) \in B_1\} \in \mathscr{B}_1 \tag{3.3.1}$$

其中 $\mathscr{B}_1$ 为 R^1 上博雷尔 σ 域，则称 $g(x)$ 是**一元博雷尔(可测)函数.**

博雷尔函数是很广泛的一类函数，我们所碰到的大部分函数都是博雷尔函数，特别地，已知所有的连续函数及单调函数都是博雷尔函数.

若 ξ 是概率空间$(\Omega,\mathscr{F},P)$ 上的随机变量，而 $g(x)$ 是一元博雷尔函数，则 $g(\xi)$ 是$(\Omega,\mathscr{F},P)$ 上的随机变量. 事实上，对一切 $B_1 \in \mathscr{B}_1$ 有

$$\{\omega : g(\xi(\omega)) \in B_1\} = \{\omega : \xi(\omega) \in g^{-1}(B_1)\} \in \mathscr{F} \tag{3.3.2}$$

这里 $g^{-1}(B_1) = \{x : g(x) \in B_1\}$，由(3.3.1) 知它是一维博雷尔点集，再根据随机变量定义中的(3.1.1) 式即得(3.3.2).

有时，我们还要考虑随机向量的函数，例如 $\eta = \xi_1 + \xi_2$ 就是随机向量(ξ_1,ξ_2) 的函数. 一般地，要研究 $\eta = g(\xi_1,\cdots,\xi_n)$ 及其分布. 这时对 n 元函数 $y = g(x_1,\cdots,x_n)$ 也要有相应的要求.

定义 3.3.2 设 $y = g(x_1,\cdots,x_n)$ 是 $\mathbf{R}^n$ 到 $\mathbf{R}^1$ 上的一个映照，若对一切 $\mathbf{R}^1$ 中的博雷尔点集 B_1 均有

$$\{(x_1,\cdots,x_n) : g(x_1,\cdots,x_n) \in B_1\} \in \mathscr{B}_n \tag{3.3.3}$$

其中 $\mathscr{B}_n$ 为 $\mathbf{R}^n$ 上博雷尔 σ 域，则称 $g(x_1,\cdots,x_n)$ 为 ***n* 元博雷尔(可测) 函数**.

同样地，若$(\xi_1,\cdots,\xi_n)$ 是$(\Omega,\mathscr{F},P)$ 上的随机向量，而 $g(x_1,\cdots,x_n)$是 n 元博雷尔函数，则 $g(\xi_1,\cdots,\xi_n)$ 是$(\Omega,\mathscr{F},P)$ 上的随机变量.

这个事实证明如下：

$$\begin{aligned} &\{\omega : g(\xi_1(\omega),\cdots,\xi_n(\omega)) \in B_1\} \\ &= \{\omega : (\xi_1(\omega),\cdots,\xi_n(\omega)) \in g^{-1}(B_1)\} \in \mathscr{F} \end{aligned} \tag{3.3.4}$$

其中 $g^{-1}(B_1)$ 是(3.3.3) 中点集，它是 n 维博雷尔点集，而按(3.2.3) 即得最后的关系式.

更一般地，还要研究 n 维随机向量$(\xi_1,\cdots,\xi_n)$ 的 m 个函数 $g_1(\xi_1,\cdots,\xi_n),\cdots,g_m(\xi_1,\cdots,\xi_n)$，这里 $g_1,\cdots,g_m$ 都是 n 元博雷尔函数，这是一个 m 维随机向量，需要求出它们的分布.

一般地来说，对于离散型随机变量，求它的函数的分布并不很困难. 例如：若 ξ 的分布列为

$$\begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

则 $\varphi(\xi)$ 的分布列可由下法得到，列出

$$\begin{pmatrix} \varphi(x_1) & \varphi(x_2) & \cdots & \varphi(x_n) & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

当然这里可能有某些 $\varphi(x_i)$ 相等，把它们作适当并项即可．

［离散褶积公式］ 若 ξ 与 η 是相互独立的随机变量，它们都取非负整数值，其概率分布分别为 $\{a_k\}$ 及 $\{b_k\}$，下面我们来计算随机变量 $\zeta = \xi + \eta$ 的概率分布．因为

$$\{\zeta = r\} = \{\xi = 0, \eta = r\} + \{\xi = 1, \eta = r - 1\} + \cdots + \{\xi = r, \eta = 0\}$$

利用独立性的假定得到

$$c_r = P\{\zeta = r\} = a_0 b_r + a_1 b_{r-1} + \cdots + a_r b_0, r = 0, 1, 2, \cdots \tag{3.3.5}$$

这就是求独立随机变量和的分布的公式 —— **离散褶积公式**，亦称**离散卷积公式**．

下面的讨论主要对连续型随机变量进行．我们将由简到繁逐步深入．

二、单个随机变量的函数的分布律

先从两个简单的例子开始．

［例 1］ 若随机变量 ξ 有密度函数 $p(x)$，而 $\eta = a\xi + b$，这里 $a \neq 0$．求 η 的密度函数 $g(y)$．

［解］ 分别记 ξ 与 η 的分布函数为 $F(x)$ 及 $G(y)$，显然有

$$G(y) = P\{\eta < y\} = P\{a\xi + b < y\}$$

$$= \begin{cases} P\left\{\xi < \dfrac{y-b}{a}\right\} = F\left(\dfrac{y-b}{a}\right), & 若\ a > 0 \\ P\left\{\xi > \dfrac{y-b}{a}\right\} = 1 - F\left(\dfrac{y-b}{a}\right), & 若\ a < 0 \end{cases} \tag{3.3.6}$$

因此

$$g(y)=\begin{cases}\dfrac{1}{a}p\left(\dfrac{y-b}{a}\right), & 若\ a>0\\ -\dfrac{1}{a}p\left(\dfrac{y-b}{a}\right), & 若\ a<0\end{cases} \tag{3.3.7}$$

或统一起来，写成

$$g(y)=\frac{1}{|a|}p\left(\frac{y-b}{a}\right) \tag{3.3.8}$$

把例 1 的结果应用到正态分布的场合．若 ξ 服从 $N(a,\sigma^2)$，则 $\eta=\dfrac{\xi-a}{\sigma}$ 的密度函数

$$g(y)=\sigma\cdot\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{\left[\sigma\left(y+\frac{a}{\sigma}\right)-a\right]^2}{2\sigma^2}\right\}=\frac{1}{\sqrt{2\pi}}e^{-y^2/2}$$

即 η 服从 $N(0,1)$，这是我们早已提到过的．

［例 2］ 若 ξ 为连续型随机变量，其分布密度为 $p(x)$，求 $\eta=\sin\xi$ 的分布．

［解］ 记 $G(y)=P\{\sin\xi<y\}$．当 $y\leqslant-1$ 时，$G(y)=0$，当 $y>1$ 时，$G(y)=1$．当 $-1<y\leqslant1$ 时（图 26），

$$\begin{aligned}G(y)&=P\{\sin\xi<y\}\\&=\sum_{k=-\infty}^{\infty}P\{2k\pi-(\pi+\sin^{-1}y)\leqslant\xi<2k\pi+\sin^{-1}y\}\\&=\int_{\sin x<y}p(x)\mathrm{d}x\end{aligned} \tag{3.3.9}$$

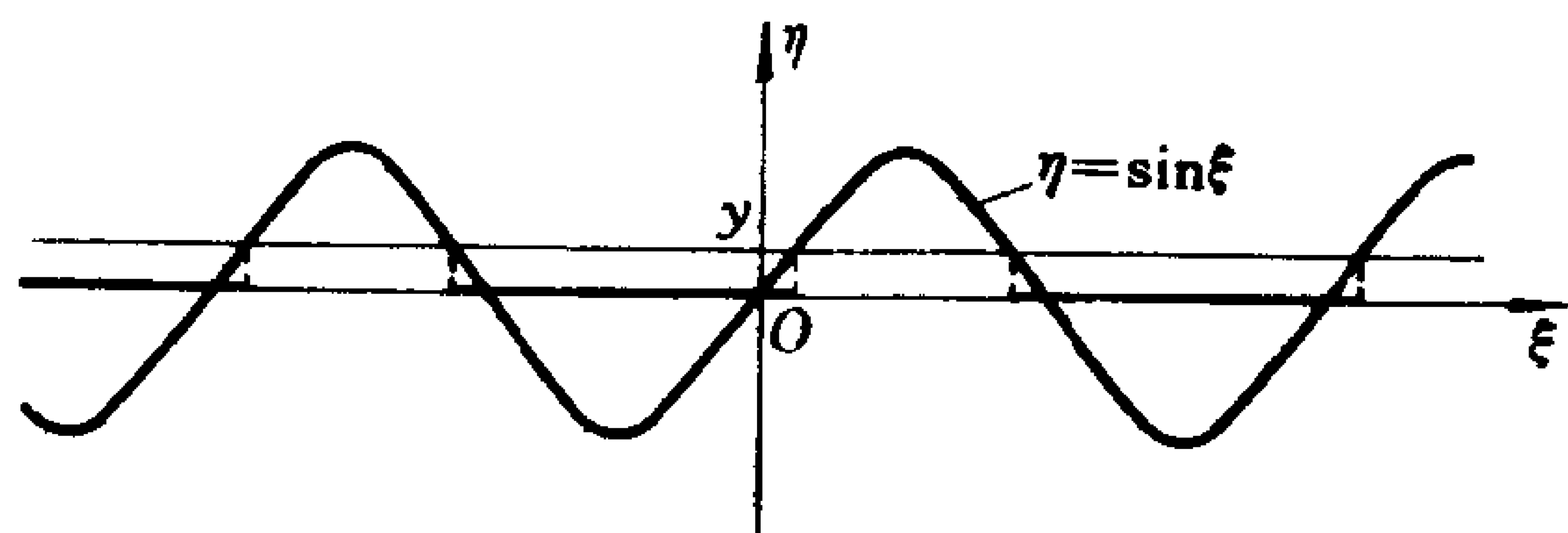


图 26 $\eta=\sin\xi$ 的分布

一般地，若 ξ 为连续型随机变量，其密度函数为 $p(x)$，而 $\eta=$

$f(\xi)$，我们对下面两种情形进行讨论.

(1) 若 $f(x)$ 严格单调，其反函数 $f^{-1}(y)$ 有连续导函数，则 $\eta = f(\xi)$ 是具有密度函数①

$$p[f^{-1}(y)]\,|[f^{-1}(y)]'| \qquad (3.3.10)$$

的连续型随机变量.

[证明] 对于任一实数 a，记使 $f(x) < a$ 成立的 x 的值的范围为 $E(a)$，则

$$\begin{aligned} P\{\eta < a\} &= P\{f(\xi) < a\} = P\{\xi \in E(a)\} \\ &= \int_{E(a)} p(x)\mathrm{d}x = \int_{-\infty}^{a} p[f^{-1}(y)]\,|[f^{-1}(y)]'|\,\mathrm{d}y \end{aligned}$$

这里用了积分的变数代换. 由上式知道 η 为连续型随机变量，其密度函数为(3.3.10).

显然可利用这个结果直接计算例 1 中的密度函数. 下面是另一个例子.

[例 3] 若 θ 服从 $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$ 的均匀分布，$\psi = \mathrm{tg}\theta$，试求 ψ 的密度函数 $g(y)$.

[解] 记 $y = \mathrm{tg}x$，则 $x = \mathrm{tg}^{-1}y$，$\frac{\mathrm{d}x}{\mathrm{d}y} = \frac{1}{1+y^2}$，因此由(3.3.10)知

$$g(y) = \frac{1}{\pi}\cdot\frac{1}{1+y^2},\ -\infty < y < \infty \qquad (3.3.11)$$

由(3.3.11)定义的分布称为**柯西分布**，它是概率论中有名的分布之一.

(2) 若 $f(x)$ 在不相重叠的区间 $I_1, I_2, \cdots$ 上逐段严格单调，其反函数分别为 $h_1(y), h_2(y), \cdots$ 而且 $h_1'(y), h_2'(y), \cdots$ 均为连续函数，那么 $\eta = f(\xi)$ 是连续型随机变量，其密度函数为

$$p[h_1(y)]\,|h_1'(y)| + p[h_2(y)]\,|h_2'(y)| + \cdots \qquad (3.3.12)$$

[证明] 给定实数 a，以 $E_i(a)$ 记 I_i 中使 $f(x) < a$ 成立的 x

① 这里和下面都约定，对使反函数无意义的 y，密度函数定义为 0.

的集合，显然诸 $E_i(a)$ 不相交，而

$$P\{\eta < a\} = P\{f(\xi) < a\} = P\left\{\xi \in \sum_i E_i(a)\right\}$$

$$= \sum_i \int_{E_i(a)} p(x)\mathrm{d}x$$

$$= \sum_i \int_{-\infty}^{a} p[h_i(y)]\,|h_i'(y)|\,\mathrm{d}y$$

$$= \int_{-\infty}^{a} \sum_i p[h_i(y)]\,|h_i'(y)|\,\mathrm{d}y$$

因此 η 是具有密度函数(3.3.12)的连续型随机变量．

这结果包含了单调的场合(3.3.10)，此外，例 2 也是其特殊场合．再看下面例子．

[例 4]　若 ξ 服从 $N(0,1)$，计算 $\eta = \xi^2$ 的密度函数．

[解]　$y = f(x) = x^2$，分段单调，在 $(-\infty,0)$ 中反函数为 $x = h_1(y) = -\sqrt{y}$，而在 $[0,\infty)$ 中反函数为 $x = h_2(y) = \sqrt{y}$，因此根据(3.3.12)知道 η 的密度函数为

$$g(y) = \varphi(-\sqrt{y})\left|-\frac{1}{2\sqrt{y}}\right| + \varphi(\sqrt{y})\left|\frac{1}{2\sqrt{y}}\right|$$

$$= \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} \mathrm{e}^{-\frac{y}{2}}, y > 0 \qquad (3.3.13)$$

顺便指出，(3.3.13) 是下列分布当 $n = 1$ 时的特例．

[χ^2 分布]　具有密度函数

$$p(x) = \frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} \mathrm{e}^{-\frac{x}{2}}, \quad x > 0 \qquad (3.3.14)$$

的分布称为**具有自由度 n 的 χ^2 分布**．

χ^2 分布在数理统计中有重要应用．与(3.1.49)中的 Γ 分布比较，我们可以知道 χ^2 分布是它的特例，其中取 $\lambda = \frac{1}{2}, r = \frac{n}{2}$，即为 $G\left(\frac{1}{2},\frac{n}{2}\right)$．以后简记 ξ 服从自由度 n 的 χ^2 分布为 $\xi \sim \chi_n^2$.

这里，我们要再提到一个有趣而重要的事实．

[均匀分布的特殊地位] 若随机变量ξ的分布函数为$F(x)$，因为$F(x)$是非降函数，对任意$0 \leqslant y \leqslant 1$，可定义

$$F^{-1}(y) = \inf\{x : F(x) > y\} \qquad (3.3.15)$$

作为$F(x)$的反函数．

下面考察随机变量$\theta = F(\xi)$的分布，这里$F(x)$是连续函数．对$0 \leqslant x \leqslant 1$，

$$\begin{aligned} P\{\theta < x\} &= P\{F(\xi) < x\} \\ &= P\{\xi < F^{-1}(x)\} = F(F^{-1}(x)) = x \end{aligned} \qquad (3.3.16)$$

即$\theta = F(\xi)$服从[0,1]均匀分布．这个结论在统计中起重要作用．

反之，若θ服从[0,1]均匀分布，对任意分布函数$F(x)$，令

$$\xi = F^{-1}(\theta) \qquad (3.3.17)$$

则

$$P\{\xi < x\} = P\{F^{-1}(\theta) < x\} = P\{\theta < F(x)\} = F(x)$$

因此ξ是服从分布函数$F(x)$的随机变量．

这样，只要我们能产生[0,1]中均匀分布的随机变量的子样（观察值），那么我们也就能通过(3.3.17)产生分布函数为$F(x)$的随机变量的子样，这结论在蒙特卡洛方法中具有基本的重要性．通常的做法是利用数学或物理的方法产生[0,1]中均匀分布随机变量的子样（称为**均匀分布随机数**），再利用变换(3.3.17)得到任意分布$F(x)$的随机数．

* **随机变量的存在性定理** 利用上述结果，我们可以给下面定理一个构造性的证明．

定理3.3.1 若$F(x)$是左连续的单调不减函数，且$F(-\infty) = 0, F(+\infty) = 1$，则存在一个概率空间$(\Omega, \mathscr{F}, P)$及其上的随机变量$\xi(\omega)$，使$\xi(\omega)$的分布函数正好是$F(x)$．

[证明] 取$\Omega = [0,1]$，再取$\mathscr{F}$为[0,1]中博雷尔点集全体，而P取为直线上的勒贝格测度（它是长度概念的推广，但对一切博雷尔点集都有定义）．定义$\theta(\omega) = \omega$，则$\theta(\omega)$是$(\Omega, \mathscr{F}, P)$上的随

机变量,又对一切 $0 \leqslant x \leqslant 1$,

$$P\{\theta(\omega) < x\} = P\{\omega \in [0,x)\} = x$$

因此 $\theta(\omega)$ 服从$[0,1]$上均匀分布.

再利用(3.3.15)定义 $F^{-1}(y)$,当然它也是单调函数,从而是博雷尔函数,令

$$\xi(\omega) = F^{-1}(\theta(\omega))$$

则 $\xi(\omega)$ 是$(\Omega,\mathscr{F},P)$ 上的随机变量,而且仿上段讨论可知,它的分布函数正好是 $F(x)$.

三、随机向量的函数的分布律

若 $\eta = f(\xi_1,\cdots,\xi_n)$,而$(\xi_1,\cdots,\xi_n)$ 的密度函数为 $p(x_1,\cdots,x_n)$,则同上面一样讨论可以得到

$$G(y) = P\{\eta < y\} = \int\cdots\int_{f(x_1,\cdots,x_n)<y} p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \tag{3.3.18}$$

我们看一些例子.

[和的分布] 若 $\eta=\xi_1+\xi_2$,而(ξ_1,ξ_2)的密度函数为 $p(x_1,x_2)$,则

$$\begin{aligned}F(y) = P\{\eta < y\} &= \iint_{x_1+x_2<y} p(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{y-x_1} p(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2\end{aligned} \tag{3.3.19}$$

特别当 ξ_1,ξ_2 相互独立时,有 $p(x_1,x_2) = p_1(x_1)p_2(x_2)$,这里 $p_1(x_1)$ 为 ξ_1 的密度函数,$p_2(x_2)$ 为 ξ_2 的密度函数.代入(3.3.19)得

$$\begin{aligned}F(y) &= \int_{-\infty}^{\infty}\mathrm{d}x_1\int_{-\infty}^{y-x_1} p_1(x_1)p_2(x_2)\mathrm{d}x_2 \\ &= \int_{-\infty}^{\infty}\mathrm{d}x_1\int_{-\infty}^{y} p_1(x_1)p_2(z-x_1)\mathrm{d}z\end{aligned}$$

$$=\int_{-\infty}^{y}\left[\int_{-\infty}^{\infty}p_1(x_1)p_2(z-x_1)\mathrm{d}x_1\right]\mathrm{d}z$$

因此η的密度函数为

$$p(x)=\int_{-\infty}^{\infty}p_1(u)p_2(x-u)\mathrm{d}u \qquad (3.3.20)$$

也可写为

$$p(x)=\int_{-\infty}^{\infty}p_1(x-u)p_2(u)\mathrm{d}u \qquad (3.3.21)$$

(3.3.20)或(3.3.21)称为**褶积公式(或卷积公式)**,在概率论中相当重要.

[商的分布] 若$\eta=\dfrac{\xi_1}{\xi_2}$,而$(\xi_1,\xi_2)$的密度函数为$p(x_1,x_2)$,则

$$G(x)=P\{\eta<x\}=P\left\{\frac{\xi_1}{\xi_2}<x\right\}=\iint_{x_1/x_2<x}p(x_1,x_2)\mathrm{d}x_1\mathrm{d}x_2$$

$$=\int_0^{\infty}\left[\int_{-\infty}^{zx}p(y,z)\mathrm{d}y\right]\mathrm{d}z+\int_{-\infty}^{0}\left[\int_{zx}^{\infty}p(y,z)\mathrm{d}y\right]\mathrm{d}z \qquad (3.3.22)$$

η的密度函数为

$$g(x)=\int_0^{\infty}p(zx,z)z\mathrm{d}z-\int_{-\infty}^{0}p(zx,z)z\mathrm{d}z$$

$$=\int_{-\infty}^{\infty}|z|p(zx,z)\mathrm{d}z \qquad (3.3.23)$$

*[关于顺序统计量的若干分布] 若$\xi_1,\xi_2,\cdots,\xi_n$是相互独立的随机变量,具有相同的分布函数$F(x)$,而ξ_n^*及ξ_1^*相当于把ξ_1,$\xi_2,\cdots,\xi_n$按大小顺序重新排列为

$$\xi_1^*\leqslant\xi_2^*\leqslant\cdots\leqslant\xi_n^* \qquad (3.3.24)$$

的末项及首项,它们在非参数统计中有重要应用.下面讨论几种与它们有关的分布.

首先求ξ_n^*的分布函数,

$$P\{\xi_n^*<x\}=P\{\max(\xi_1,\xi_2,\cdots,\xi_n)<x\}$$

$$=P\{\xi_1<x,\xi_2<x,\cdots,\xi_n<x\}$$
$$=P\{\xi_1<x\}\cdot P\{\xi_2<x\}\cdots P\{\xi_n<x\}$$
$$=[F(x)]^n \tag{3.3.25}$$

其次求 ξ_1^* 的分布函数,注意到

$$P\{\xi_1^*\geqslant x\}=P\{\min(\xi_1,\xi_2,\cdots,\xi_n)\geqslant x\}$$
$$=P\{\xi_1\geqslant x,\xi_2\geqslant x,\cdots,\xi_n\geqslant x\}$$
$$=P\{\xi_1\geqslant x\}P\{\xi_2\geqslant x\}\cdots P\{\xi_n\geqslant x\}$$
$$=[1-F(x)]^n$$

因此

$$P\{\xi_1^*<x\}=1-[1-F(x)]^n \tag{3.3.26}$$

进一步,讨论(ξ_1^*,ξ_n^*)的联合分布.

记$G(x,y)=P\{\xi_1^*<x,\xi_n^*<y\}$.

若 $x\geqslant y$,则

$$G(x,y)=P\{\xi_1^*<x,\xi_n^*<y\}$$
$$=P\{\xi_n^*<y\}=[F(y)]^n \tag{3.3.27}$$

若 $x<y$,则

$$G(x,y)=P\{\xi_1^*<x,\xi_n^*<y\}$$
$$=P\{\xi_n^*<y\}-P\{\xi_1^*\geqslant x,\xi_n^*<y\}$$
$$=[F(y)]^n-[F(y)-F(x)]^n \tag{3.3.28}$$

其联合密度函数为

$$g(x,y)=\begin{cases}0, & x\geqslant y\\ n(n-1)[F(y)-F(x)]^{n-2}f(x)f(y), & x<y\end{cases} \tag{3.3.29}$$

最后,我们来求**极差** $R=\xi_n^*-\xi_1^*$ 的分布密度函数 $f_R(r)$,显然对 $r\leqslant 0$,$f_R(r)=0$,若 $r>0$,则

$$P\{R<r\}=\iint\limits_{y-x<r}g(x,y)\mathrm{d}x\mathrm{d}y$$
$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x+r}g(x,y)\mathrm{d}y\right]\mathrm{d}x$$

$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{r}g(x,x+z)\mathrm{d}z\right]\mathrm{d}x$$

$$=\int_{-\infty}^{r}\left[\int_{-\infty}^{\infty}g(x,x+z)\mathrm{d}x\right]\mathrm{d}z$$

因此

$$f_R(x)=\int_{-\infty}^{\infty}g(x,x+r)\mathrm{d}x$$

$$=n(n-1)\int_{-\infty}^{\infty}[F(x+r)-F(x)]^{n-2}\cdot f(x)f(x+r)\mathrm{d}x \qquad (3.3.30)$$

四、随机向量的变换

若$(\xi_1,\cdots,\xi_n)$的密度函数为$p(x_1,\cdots,x_n)$，求$\eta_1=f_1(\xi_1,\cdots,\xi_n),\cdots,\eta_m=f_m(\xi_1,\cdots,\xi_n)$的分布. 这时有

$$G(y_1,\cdots,y_m)=P\{\eta_1<y_1,\cdots,\eta_m<y_m\}$$

$$=\int\cdots\int_{\substack{f_1(x_1,\cdots,x_n)<y_1\\ \cdots\\ f_m(x_1,\cdots,x_n)<y_m}}p(x_1,\cdots,x_n)\mathrm{d}x_1\cdots\mathrm{d}x_n \qquad (3.3.31)$$

显然，这是最一般的场合．当$m=1$时便是随机向量的函数的情形，当$m=n=1$时得到单个随机变量的函数的情形．下面考虑另一个重要的特殊情形，即当$(\xi_1,\cdots,\xi_n)$与$(\eta_1,\cdots,\eta_m)$有一一对应变换关系时，当然这时$n=m$必须成立．

若对$y_i=f_i(x_1,\cdots,x_n)$存在唯一的反函数$x_i(y_1,\cdots,y_n)=x_i(i=1,\cdots,n)$，且$(\eta_1,\cdots,\eta_n)$的密度函数为$q(y_1,\cdots,y_n)$，那么

$$G(y_1,\cdots,y_n)=\int\cdots\int_{\substack{u_1<y_1\\ \cdots\\ u_n<y_n}}q(u_1,\cdots,u_n)\mathrm{d}u_1\cdots\mathrm{d}u_n \qquad (3.3.32)$$

比较(3.3.31)与(3.3.32)可知

$q(y_1,\cdots,y_n)$

$$= \begin{cases} p(x_1,\cdots,x_n)\,|J|, & 若(y_1,\cdots,y_n)属于 f_1,\cdots,f_n 的值域 \\ 0, & 其他 \end{cases}$$

(3.3.33)

其中 J 为坐标变换的雅可比行列式

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \cdots & \cdots & \cdots \\ \frac{\partial x_1}{\partial y_n} & \cdots & \frac{\partial x_n}{\partial y_n} \end{vmatrix} \qquad (3.3.34)$$

这里，我们假定上述偏导数存在而且连续.

公式（3.3.33）对应于单变量场合的公式（3.3.10），也可以导出对应于公式（3.3.12）的多变量场合的公式，这留给读者作为练习.

［例 5］ 若（ξ_1，ξ_2）的密度函数为 p（x_1，x_2），而

$$\begin{aligned} \eta_1 &= a\xi_1 + b\xi_2 \\ \eta_2 &= c\xi_1 + d\xi_2 \end{aligned} \qquad (3.3.35)$$

这里 $\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$，试求$(\eta_1,\eta_2)$的密度函数 $q(y_1,y_2)$.

［解］ 在本例中

$$\begin{aligned} y_1 &= f_1(x_1,x_2) = ax_1 + bx_2, \\ y_2 &= f_2(x_1,x_2) = cx_1 + dx_2 \end{aligned}$$

因此

$$x_1 = \frac{d}{\Delta}y_1 - \frac{b}{\Delta}y_2, \quad x_2 = -\frac{c}{\Delta}y_1 + \frac{a}{\Delta}y_2$$

而

$$J = \begin{vmatrix} \frac{d}{\Delta} & -\frac{b}{\Delta} \\ -\frac{c}{\Delta} & \frac{a}{\Delta} \end{vmatrix} = \frac{ad-bc}{\Delta^2} = \frac{1}{ad-bc}$$

最后得到

$$q(y_1,y_2)=\frac{p\left(\frac{d}{\Delta}y_1-\frac{b}{\Delta}y_2,\ -\frac{c}{\Delta}y_1+\frac{a}{\Delta}y_2\right)}{|ad-bc|}$$

(3.3.36)

［例 6］　若ξ，η为独立随机变量，且分别服从$G(\lambda,r_1)$及$G(\lambda,r_2)$，试求$\alpha=\xi+\eta$与$\beta=\frac{\xi}{\xi+\eta}$的密度函数$q(u,v)$.

［解］　对$u\geqslant 0$，$v\geqslant 0$作变换$u=x+y$，$v=\frac{x}{x+y}$，因此$x=uv$，$y=u(1-v)$，$0<v<1$.

$$J=\begin{vmatrix}\frac{\partial x}{\partial u} & \frac{\partial x}{\partial v}\\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}v & u\\ 1-v & -u\end{vmatrix}=-u$$

所以

$$|J|=u,\qquad u\geqslant 0$$

这样，α，β的联合密度函数为

$$\begin{aligned}q(u,v)&=\frac{\lambda^{r_1}}{\Gamma(r_1)}x^{r_1-1}e^{-\lambda x}\cdot\frac{\lambda^{r_2}}{\Gamma(r_2)}y^{r_2-1}e^{-\lambda y}|J|\\&=\frac{\lambda^{r_1+r_2}}{\Gamma(r_1)\Gamma(r_2)}(uv)^{r_1-1}[u(1-v)]^{r_2-1}e^{-\lambda u}\cdot u\\&=\frac{\lambda^{r_1+r_2}}{\Gamma(r_1+r_2)}u^{r_1+r_2-1}e^{-\lambda u}\\&\quad\cdot\frac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)}v^{r_1-1}(1-v)^{r_2-1},\quad u\geqslant 0,0<v<1\end{aligned}$$

因此，α与β独立，而且α服从$G(\lambda,r_1+r_2)$，这揭示了Γ分布的一个重要性质：若两个相互独立的Γ分布随机变量ξ与η，具有相同的尺度参数λ，各具有形状参数r_1及r_2，则其和$\xi+\eta$服从$G(\lambda,r_1+r_2)$. 这个性质称为Γ分布的再生性，简记作$G(\lambda,r_1)*G(\lambda,r_2)=G(\lambda,r_1+r_2)$.

随机变量β服从如下分布

$$
p(v)=\begin{cases}\dfrac{\Gamma(r_1+r_2)}{\Gamma(r_1)\Gamma(r_2)}v^{r_1-1}(1-v)^{r_2-1}, & 0<v<1\\ 0, & \text{其他}\end{cases}
\tag{3.3.37}
$$

这个分布称为**β分布**，r_1 及 r_2 为它的参数，这也是数理统计中的一个重要分布.

五、随机变量的函数的独立性

首先证明一个定理.

定理 3.3.2 若 ξ_1，…，ξ_n 是相互独立的随机变量，则 $f_1(\xi_1)$，…，$f_n(\xi_n)$也是相互独立的，这里 f_i ($i=1$，…，n) 是任意的一元博雷尔函数.

［证明］ 对任意的一维博雷尔点集 $A_1,\cdots,A_n$ 有

$$
\begin{aligned}
&P\{f_1(\xi_1)\in A_1,\cdots,f_n(\xi_n)\in A_n\}\\
=&P\{\xi_1\in f_1^{-1}(A_1),\cdots,\xi_n\in f_n^{-1}(A_n)\}\\
=&P\{\xi_1\in f_1^{-1}(A_1)\}\cdots P\{\xi_n\in f_n^{-1}(A_n)\}\\
=&P\{f_1(\xi_1)\in A_1\}\cdots P\{f_n(\xi_n)\in A_n\}
\end{aligned}
$$

定理的结论在直观上是明显的，但在定理的证明中却要两次用到未证明的论断(3.2.37)其中第一次用来指明对 $\xi_1,\cdots,\xi_n$ 的有关概率可以化为乘积的形式，另一次用来说明最后的等式表明 $f_1(\xi_1),\cdots,f_n(\xi_n)$是相互独立的，且第一次是难以避免的.

这个结果可以推广到随机向量的场合.

例 6 说明，即使由相同的随机向量构成的不同函数也可能是独立的，这种情况在概率论与数理统计中相当重要，下面再讨论一些例子.

［例 7］ 若 ξ 与 η 是相互独立的随机变量，均服从 $N(0,1)$，试证化为极坐标后，$\rho=\sqrt{\xi^2+\eta^2}$ 及 $\varphi=\operatorname{arctg}\left(\dfrac{\eta}{\xi}\right)$，取值于 $[0,2\pi]$，是相互独立的.

［解］ 采用极坐标，$x=r\cos\theta$，$y=r\sin\theta$，因此 $r=\sqrt{x^2+y^2}$，$\theta=\mathrm{arctg}\,\frac{y}{x}$，因为（$\xi$，$\eta$）的密度函数为

$$p(x,y)=\frac{1}{2\pi}e^{-(x^2+y^2)/2}$$

而

$$J=\begin{vmatrix}\frac{\partial x}{\partial r} & \frac{\partial x}{\partial\theta}\\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial\theta}\end{vmatrix}=\begin{vmatrix}\cos\theta & -r\sin\theta\\ \sin\theta & r\cos\theta\end{vmatrix}=r$$

故（ρ，φ）的密度函数为

$$q(r,\theta)=\frac{1}{2\pi}e^{-(x^2+y^2)/2}\cdot r$$

$$=\frac{1}{2\pi}\cdot re^{-r^2/2},\quad r\geqslant 0,\quad 0\leqslant\theta\leqslant 2\pi \tag{3.3.38}$$

即 $\rho=\sqrt{\xi^2+\eta^2}$的密度函数为

$$R(r)=\begin{cases}re^{-r^2/2}, & r\geqslant 0\\ 0, & r<0\end{cases} \tag{3.3.39}$$

这个分布称为**瑞利**（Rayleigh）**分布**.

而 $\varphi=\mathrm{arctg}\,\frac{\eta}{\xi}$服从［0，2π］中的均匀分布，并且它们是独立的.

这个结果常被用来产生服从正态分布的随机数.做法如下：产生相互独立的［0，1］均匀分布的随机数 U_1，U_2，令

$$\xi=(-2\ln U_1)^{\frac{1}{2}}\cos 2\pi U_2$$

$$\eta=(-2\ln U_1)^{\frac{1}{2}}\sin 2\pi U_2 \tag{3.3.40}$$

则 ξ 与 η 是相互独立的 N（0，1）随机数.

这样做法的理由让读者自行论证.

［例 8］ 若（ξ_1，ξ_2）服从二元正态分布（3.2.22），其中 $a_1=0$，$a_2=0$. 令

$$\eta_1=\xi_1\cos\alpha+\xi_2\sin\alpha,\quad \eta_2=-\xi_1\sin\alpha+\xi_2\cos\alpha \tag{3.3.41}$$

这里 $0\leqslant\alpha\leqslant 2\pi$，是某个角度. 我们来求 (η_1, η_2) 的密度函数 $q(u, v)$.

这里可直接用例 5 的结果. 其中 $J=1$，因此

$$q(u, v) = p\ (u\cos\alpha - v\sin\alpha,\ u\sin\alpha + v\cos\alpha)$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}}\exp\left\{-\frac{1}{2(1-r^2)}(Au^2-2Buv+Cv^2)\right\} \tag{3.3.42}$$

其中

$$A=\frac{\cos^2\alpha}{\sigma_1^2}-2r\frac{\cos\alpha\sin\alpha}{\sigma_1\sigma_2}+\frac{\sin^2\alpha}{\sigma_2^2}$$

$$B=\frac{\cos\alpha\sin\alpha}{\sigma_1^2}-r\frac{\sin^2\alpha-\cos^2\alpha}{\sigma_1\sigma_2}-\frac{\cos\alpha\sin\alpha}{\sigma_2^2}$$

$$C=\frac{\sin^2\alpha}{\sigma_1^2}+2r\frac{\cos\alpha\sin\alpha}{\sigma_1\sigma_2}+\frac{\cos^2\alpha}{\sigma_2^2}$$

由 (3.3.42) 可看出由二维正态变量 (ξ_1, ξ_2) 经坐标旋转而得的随机向量 (η_1, η_2) 还是服从正态分布. 进一步，若选 α 使得

$$\mathrm{tg}2\alpha=\frac{2r\sigma_1\sigma_2}{\sigma_1^2-\sigma_2^2} \tag{3.3.43}$$

则 $B=0$，因此 η_1 与 η_2 独立. 这说明二元正态分布可经适当的坐标旋转化为两个独立正态分布之积. 利用正交变换把多元正态分布化作独立正态分布之积，在数理统计中有重要应用.

第三章小结

本章中我们详细地研究了随机变量. 用随机变量描述随机现象是近代概率论中最重要的方法，以后我们所讨论的随机事件几乎都用随机变量来描述.

我们给出了随机变量的严格定义，按照这种观点，随机变量是定义在样本空间上的具有某种可测性的实值函数，只是出于历史的原因，才沿用“变量”二字.

对于随机变量，重要的是要知道它取哪一些值以及以怎样的

概率取这些值．从这个角度讲，分布函数完整地描述了随机变量．同时，分布函数具有良好的分析性质，很便于研究，它是研究随机变量的重要工具．

离散型随机变量与连续型随机变量是最重要的两类随机变量，由于它们的取值特征不同，因此对它们的描述及处理方法都有很大不同，应该进行对比，从而加深理解．

我们还介绍了多种分布函数，而且比较注重各种分布函数的特征性质及它们之间的联系．这些分布函数在不同的理论和实际问题中扮演着重要角色．

正态分布是概率论中最重要的分布，在应用中及理论研究中占有头等重要的地位，它与二项分布及泊松分布是概率论中最重要的三种分布．判断一种分布重要性的标准是：(1) 在实际工作中经常遇到；(2) 在理论研究中重要，具有较好的性质；(3) 用它能导出许多重要分布．以上三种分布都满足这些要求．

把几个随机变量放在一起研究时不但需要研究各个分量个别的性质，而且要考虑它们之间的联系，从而大大丰富了研究的内容，条件分布及独立性概念也随之出现，它们是条件概率及事件独立性概念在随机变量场合的具体化，在今后研究中很重要．

随机变量的函数的分布律的推导，在数理统计中及在概率论的许多应用中相当重要．我们这里分三种类型介绍了最基本的处理方法．这部分内容，只有通过动手多作练习，才能牢固掌握．

习 题 三

1. 直线上有一质点，每经一个单位时间，它分别以概率 p 及 q 向右或向左移动一格，若该质点在时刻 0 从原点出发，而且每次移动是相互独立的，试用随机变量来描述这质点的运动（以 S_n 表示时刻 n 时质点的位置）．

2. 设 ξ 为伯努利试验中第一个**游程**（连续的成功或失败）的长，试求 ξ 的概率分布．

3. C 应取何值才能使下列函数成为概率分布：

(1) $f(k)=\frac{C}{N}$，$k=1, 2, \cdots, N$；

(2) $f(k)=C\frac{\lambda^k}{k!}$，$k=1, 2, \cdots$，$\lambda>0$.

4. 若分布函数定义为 $F(x)=P\{\xi\leqslant x\}$，试证这时的 $F(x)$ 具有下列性质：(i) 非降；(ii) $F(-\infty)=0$，$F(+\infty)=1$；(iii) 右连续.

5. 证明函数

$$f(x)=\frac{1}{2\lambda}e^{-\frac{|x-\mu|}{\lambda}},\ -\infty<x<\infty,\quad \lambda>0,\ \mu \text{常数}$$

是一个密度函数. 这个分布称为**拉普拉斯分布**.

6. 妊娠天数 ξ 的分布函数为 $N(270, 100)$，求 ξ 落在下列范围的概率：(i) (260, 280)；(ii) 短于 250 天；(iii) 长于 300 天.

7. 若 ξ 的分布函数为 $N(60, 9)$，求分点 x_1，x_2，x_3，x_4，使 ξ 落在 $(-\infty, x_1)$，(x_1, x_2)，(x_2, x_3)，(x_3, x_4)，(x_4, ∞) 中的概率之比为 7 : 24 : 38 : 24 : 7.

*8. 在巴斯卡分布 $\binom{k-1}{r-1}p^rq^{k-r}$ 中，令 $k\Delta t=t$，$p=\lambda\Delta t$，试证当 $\Delta t\to 0$ 时，它能用 $\frac{\lambda(\lambda t)^{r-1}}{(r-1)!}e^{-\lambda t}\cdot\Delta t$ 来逼近.（这可以解释为第 r 次成功发生在 $(t, t+\Delta t)$ 中的概率，其密度函数正好是参数为 r 的埃尔兰分布. 用这种方法可以把课文中的对比严格化.）

9. 在生存分析中，作为研究对象的是非负随机变量，它们的分布称为寿命分布. 若 ξ 是非负随机变量，其分布函数为 $F(x)$，密度函数为 $f(x)$，这时通常还引入**生存函数** $S(x)=P\{\xi\geqslant x\}$ 及**失效率函数** $\lambda(x)=\frac{f(x)}{1-F(x)}$. 试导出 $S(x)$，$\lambda(x)$，$F(x)$ 及 $f(x)$ 之间的关系式.

10. 设随机变量 ξ 取值于 [0, 1]，若 $P\{x\leqslant\xi<y\}$ 只与长度 $y-x$ 有关（对一切 $0\leqslant x\leqslant y\leqslant 1$），试证 ξ 服从 [0, 1] 均匀分布.

*11. 若存在 Θ 上的实值函数 $Q(\theta)$ 及 $D(\theta)$ 以及 $T(x)$ 及 $S(x)$，使

$$f_\theta(x)=\exp\{Q(\theta)T(x)+D(\theta)+S(x)\}$$

则称 $\{f_\theta, \theta\in\Theta\}$ 是一个**单参数的指数族**. 证明 (1) 正态分布 $N(m_0, \sigma^2)$，已知 m_0，关于参数 σ；(2) 正态分布 $N(m, \sigma_0^2)$，已知 σ_0，关于参数 m；(3) 泊松分布 $P(\lambda)$ 关于 λ 是一个单参数的指数族.

但是 $[0, \theta]$ 上均匀分布，关于 θ 不是一个单参数的指数族.

12. 定义二元函数

$$F(x,y)=\begin{cases}1, & x+y>0\\ 0, & x+y\leqslant 0\end{cases}$$

验证此函数对每个变元非降，左连续，且满足分布函数性质（ii)，但无法使（3.2.5）保持非负.

*13. 试证 $f(x, y)=Ke^{-(ax^2+2bxy+cy^2)}$ 为密度函数的充要条件为 $a>0$，$c>0$，$b^2-ac<0$，$K=\frac{1}{\pi}\sqrt{ac-b^2}$.

14. 若 $f_1(x)$，$f_2(y)$ 为分布密度，求为使 $f(x, y)=f_1(x)\times f_2(y)+h(x,y)$ 成为密度函数，$h(x,y)$ 必须而且只须满足什么条件.

15. 若 (ξ, η) 的密度函数为

$$f(x,y)=\begin{cases}Ae^{-(2x+y)}, & x>0, y>0\\ 0, & \text{其他}\end{cases}$$

试求：(1) 常数 A；(2) $P\{\xi<2, \eta<1\}$；(3) ξ 的边际分布；(4) $P\{\xi+\eta<2\}$；(5) $f(x|y)$；(6) $P\{\xi<2|\eta<1\}$.

*16. 若 (ξ, η) 服从二元正态分布，参数 a_1，a_2，σ_1^2，σ_2^2，r，以 $D(\lambda)$ 记下面椭圆的内部：

$$\frac{(x-a_1)^2}{\sigma_1^2}-\frac{2r(x-a_1)(y-a_2)}{\sigma_1\sigma_2}+\frac{(y-a_2)^2}{\sigma_2^2}=\lambda^2$$

试求 $P\{(\xi, \eta)\in D(\lambda)\}$.

17. 证明多项分布的边际分布仍为多项分布.

18. 设二维随机变量 (ξ, η) 的联合密度为

$$p(x,y)=\frac{1}{\Gamma(k_1)\Gamma(k_2)}x^{k_1-1}(y-x)^{k_2-1}e^{-y}$$

$k_1>0$，$k_2>0$，$0<x\leqslant y<\infty$. 试求 ξ 与 η 的边际分布.

19. 若 $f_1(x)$，$f_2(x)$，$f_3(x)$ 是对应于分布函数 $F_1(x)$，$F_2(x)$，$F_3(x)$ 的密度函数，证明对于一切 α $(-1<\alpha<1)$，下列函数是密度函数，且具有相同的边际密度函数 $f_1(x)$，$f_2(x)$，$f_3(x)$：

$$\begin{aligned}&f_\alpha(x_1, x_2, x_3)\\ &=f_1(x_1)f_2(x_2)f_3(x_3)\{1+\alpha[2F_1(x_1)-1]\\ &\quad\cdot[2F_2(x_2)-1][2F_3(x_3)-1]\}\end{aligned}$$

20. 若每次试验中出现 A_1，A_2，A_3 的概率分别为 p_1，p_2 及 p_3，而且 $p_1+p_2+p_3=1$，共进行 n 次独立试验，以 ξ_i 记 A_i 出现的次数，试求 $P\{\xi_1=k|\xi_2=m\}$.

21. (1) 若 (ξ, η) 的联合密度函数为

$$f(x,y)=\begin{cases}4xy, 0\leqslant x\leqslant 1,\ 0\leqslant y\leqslant 1\\ 0, \quad \text{其他}\end{cases}$$

问 ξ 与 η 是否相互独立？

(2) 若 $(\xi,\ \eta)$ 的联合密度函数为

$$g(x,y)=\begin{cases}8xy, & 0\leqslant x\leqslant y, 0\leqslant y\leqslant 1\\ 0, & \text{其他}\end{cases}$$

问 ξ 与 η 是否相互独立？

22. 设 $(\xi,\ \eta,\ \zeta)$ 的联合密度函数为

$$p(x,y,z)=\begin{cases}\dfrac{1}{8\pi^3}(1-\sin x\sin y\sin z), & \text{当}\begin{array}{l}0\leqslant x\leqslant 2\pi\\0\leqslant y\leqslant 2\pi\\0\leqslant z\leqslant 2\pi\end{array}\text{时}\\ 0, & \text{其他}\end{cases}$$

试证 ξ，η，ζ 两两独立，但不相互独立.

23. 设 $(\xi,\ \eta)$ 具有联合密度函数

$$p(x,y)=\begin{cases}\dfrac{1+xy}{4}, & |x|<1, |y|<1\\ 0, & \text{其他}\end{cases}$$

试证 ξ 与 η 不独立，但 ξ^2 与 η^2 是相互独立的.

24. 若 ξ，η 相互独立且皆以概率 $\frac{1}{2}$ 取值 $+1$ 及 -1，令 $\zeta=\xi\eta$，试证 ξ，η，ζ 两两独立但不相互独立.

25. 若 ξ_1 与 ξ_2 是独立随机变量，且 $\xi_1\sim B(n_1,\ p)$，$\xi_2\sim B(n_2,\ p)$，试直接证明

(1) $\xi_1+\xi_2\sim B(n_1+n_2,\ p)$；

(2) $P\{\xi_1=k|\xi_1+\xi_2=n\}=\dfrac{\dbinom{n_1}{k}\dbinom{n_2}{n-k}}{\dbinom{n_1+n_2}{n}}$.

26. 若 ξ_1 与 ξ_2 是独立随机变量，均服从泊松分布，参数分别为 λ_1 及 λ_2，试直接证明：

(1) $\xi_1+\xi_2$ 具有泊松分布，参数为 $\lambda_1+\lambda_2$；

(2) $P\{\xi_1=k|\xi_1+\xi_2=n\}=\dbinom{n}{k}\left(\dfrac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\dfrac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k}$.

27. 设 ξ 的密度函数为 $p(x)$，求下列随机变量的分布函数：(1) $\eta=\xi^{-1}$；(2) $\eta=\mathrm{tg}\xi$；(3) $\eta=|\xi|$.

28. 设 ξ 与 η 相互独立且服从同一几何分布，令 $\zeta=\max(\xi,\ \eta)$，试求(1) $(\zeta,\ \xi)$ 的联合分布；(2) ζ 的分布；(3) ξ 关于 ζ 的条件概率分布.

29. 若 ξ，η 为相互独立的分别服从 [0. 1] 均匀分布的随机变量，试求 $\zeta=\xi+\eta$ 的分布密度函数.

30. 设 ξ，η 相互独立，分别服从 N (0，1)，试求 $\zeta=\dfrac{\xi}{\eta}$ 的密度函数.

31. 若 ξ，η 是独立随机变量，均服从 N (0，1)，试求 $U=\xi+\eta$，$V=\xi-\eta$ 的联合密度函数.

32. 若 ξ_1，ξ_2，…，ξ_n 相互独立，且皆服从指数分布，参数分别为 λ_1，λ_2，…，λ_n，试求 $\eta=\min$ (ξ_1，ξ_2，…，ξ_n) 的分布.

33. 在 (0，a) 线段上随机投掷两点，试求两点间距离的分布函数.

34. 若气体分子的速度是随机向量 $V=$ (X，Y，Z)，各分量相互独立，且均服从 N (0，σ^2)，试证 $S=\sqrt{X^2+Y^2+Z^2}$ 服从**马克斯威尔分布**

$$f(s)=\sqrt{\frac{2}{\pi}}\frac{s^2}{\sigma^3}\exp\left(-\frac{s^2}{2\sigma^2}\right),\quad s>0$$

35. 设 X 是取正值的随机变量，若 $\ln X\sim N$ (μ，σ^2)，试证 X 的密度函数为

$$p(x)=\begin{cases}\dfrac{1}{\sigma x\sqrt{2\pi}}\exp\left\{-\dfrac{1}{2\sigma^2}(\ln x-\mu)^2\right\}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

这个分布称为**对数正态分布**.

36. 通称下列分布函数为**韦布尔分布**：

$$F(x)=\begin{cases}1-e^{-\lambda x^{\alpha}}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

这是韦布尔 (Weibull) 在研究金属材料的疲劳寿命中导出的，在可靠性研究中有广泛应用.

若 ξ_1，ξ_2，…，ξ_n 相互独立，相同分布，并以 ξ_1^* 记它们的最小值，则当 $\xi_i\sim F$ (x) 时，试求 ξ_1^* 的分布函数. 反过来，若 $\xi_1^*\sim F$ (x)，试导出 ξ_i 的分布函数.

链条的寿命取决于最弱环节，试说明上述概率结论的实际含意.

*37. 设 ξ，η 为两个独立随机变量，ξ 服从 N (0，1)，η 服从自由度为 n 的 χ^2 分布 (3.3.14)，令 $t=\xi\Big/\sqrt{\dfrac{\eta}{n}}$，试证 t 的密度函数为

$$p_n(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)}\left(\frac{x^2}{n}+1\right)^{-(n+1)/2}$$

这分布称为**具有自由度 n 的 t 分布**，在数理统计中十分重要.

*38. 若 ξ 与 η 独立，且 $\xi\sim\chi_m^2$，$\eta\sim\chi_n^2$. 试证：

(1) $\xi+\eta$ 与 $\frac{\xi}{\eta}\cdot\frac{n}{m}$ 独立；

(2) $\frac{\xi}{\eta}\frac{n}{m}$ 的密度函数为

$$p(x)=\begin{cases}\dfrac{\Gamma\left(\dfrac{m+n}{2}\right)}{\Gamma\left(\dfrac{m}{2}\right)\Gamma\left(\dfrac{n}{2}\right)}\dfrac{\left(\dfrac{m}{n}\right)^{\frac{m}{2}}x^{\frac{m}{2}-1}}{\left(1+\dfrac{m}{n}x\right)^{\frac{m+n}{2}}}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

这个分布称为 **F 分布**.

39. 若 ξ，η 独立，且均服从 N（0，1），试证 $U=\xi^2+\eta^2$ 与 $V=\frac{\xi}{\eta}$ 是独立的，并求出它们的密度函数.

40. 若（ξ，η）服从二元正态分布（3.2.22），试找出 $\xi+\eta$ 与 $\xi-\eta$ 相互独立的充要条件.

41. 对二元正态密度函数

$$p(x,y)=\frac{1}{2\pi}\exp\left\{-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)\right\}$$

(1) 把它化为标准形式（3.2.22）；(2) 指出 a_1，a_2，σ_1，σ_2，r；(3) 求 p_1（x）；(4) 求 P（$x|y$）.

42. 设 $\boldsymbol{a}=\boldsymbol{0}$，$\boldsymbol{B}^{-1}=\begin{pmatrix}7&3&2\\3&4&1\\2&1&2\end{pmatrix}$，试写出分布密度（3.2.12）并求出（$\xi_1$，$\xi_2$）的边际密度函数.

**43. 设 ξ 与 η 是相互独立相同分布的随机变量，其密度函数不等于 0 且有二阶导数，试证若 $\xi+\eta$ 与 $\xi-\eta$ 相互独立，则随机变量 ξ，η，$\xi+\eta$，$\xi-\eta$ 均服从正态分布.

44. 把习题 25 和习题 26 的结果推广到 n 个随机变量的场合.

45. 设随机变量 ξ 与 η 相互独立，$\xi\sim B$（n，p），而 η 服从（0，1）上均匀分布，试求 $\xi+\eta$ 的分布函数和密度函数.

**46. 试求顺序统计量 ξ_k^* 与 ξ_l^*（$k<l$）的联合密度函数.

47. 试利用概率的连续性重新证明一元分布函数的性质（ii）和性质（iii），并说明这种证法可推广到多元的场合.

*48. 利用随机变量分布解释贝特朗奇论.

*49. 若 f 是 Ω 上单值实函数，对 $B\subset \mathbf{R}^1$，记 $f^{-1}(B)=\{\omega\in\Omega: f(\omega)\in B\}$，试证逆映照 f^{-1} 具有如下性质：

(1) $f^{-1}(\bigcup_{\lambda\in\Lambda}B_\lambda)=\bigcup_{\lambda\in\Lambda}f^{-1}(B_\lambda)$；

(2) $f^{-1}(\bigcap_{\lambda\in\Lambda}B_\lambda)=\bigcap_{\lambda\in\Lambda}f^{-1}(B_\lambda)$；

(3) $f^{-1}(\overline{B})=\overline{f^{-1}(B)}$.

**50. 证明：ξ 是一个随机变量当且仅当对任何 $x\in\mathbf{R}^1$，成立

$$\{\omega: \xi(\omega)<x\}\in\mathscr{F}$$

[提示：必要性是显然的，为证充分性，记 $\mathbf{m}=\{A: A\subset\mathbf{R}^1, (\xi(\omega)\in A)\in\mathscr{F}\}$，验证 $\mathbf{m}$ 是 σ 域，又 $\mathbf{m}$ 包含全体形如 $(-\infty, x)$ 的区间，故 $\mathbf{m}$ 包含 $\mathscr{F}_1$.]

第四章　数字特征与特征函数

§1. 数学期望

一、平均值与加权平均值

有甲、乙两个射手，他们的射击技术用下表表出：

甲射手

击中环数	8	9	10
概　　率	0.3	0.1	0.6

乙射手

击中环数	8	9	10
概　　率	0.2	0.5	0.3

试问哪一个射手技术较好？

这个问题的答案不是一眼看得出的. 这说明分布列虽然完整地描述了随机变量，但是却不够“集中”地反映出它的变化情况. 因此我们有必要找出一些量来更集中、更概括地描述随机变量，这些量多是某种平均值.

求平均值是大家都很熟悉的一种运算. 例如，某公司有 n 个职工，他们的工资分别为 $x_1, x_2, \cdots, x_n$，则这个公司的平均工资为

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \qquad (4.1.1)$$

还有其他求平均值的方法. 例如，一个小学生，他的考试成绩为：语文 95 分，算术 85 分，常识 60 分，若依上面方法计算，则他的平均成绩为 $\bar{x}=80$. 显然，这个数字不太能反映这个学生的真正成绩，因为它没有考虑到这三个科目的相对重要性. 在这个年级中，每周语文有 10 节课，数学有 8 节，而常识只有 2 节. 在

评价学生成绩时，这个因素不能不考虑，因此用下面方法来计算学生的平均成绩，似乎更合理些：

$$\bar{x}_w = \frac{95\times 10 + 85\times 8 + 60\times 2}{10+8+2}$$

$$= 95\times\frac{10}{20} + 85\times\frac{8}{20} + 60\times\frac{2}{20} = 87.5$$

这种平均称为加权平均. 其一般定义如下：给定权 $w_i\geqslant 0$，$i=1, 2, \cdots, n$，满足 $\sum_{i=1}^{n} w_i = 1$，则

$$\bar{x}_w = \sum_{i=1}^{n} w_i x_i \qquad (4.1.2)$$

称为 $x_1, x_2, \cdots, x_n$ 关于权 $\{w_i, i=1, 2, \cdots, n\}$ 的**加权平均值**.

显然，在某些情况下，加权平均更加合理. 由于“权”的大小直接影响最后结果，因此“权”的选择是加权平均中最重要的问题. 例如为测量泰山的高度可以安排几个测量队从不同的地点进行测量，由于各个测量队的技术水平不一，各地点的地理条件不同，因此最后的结果恐怕要采用某种加权平均值为好.

不言而喻，普通平均是加权平均的一种特例，这时所有的权相等.

平均值按其大小总在原始数据的当中，因此它反映了一组数据的中心趋势.

在上面的问题中，若使两个射手各射 N 枪，则他们打中的环数大约是：

甲：$8\times 0.3N + 9\times 0.1N + 10\times 0.6N = 9.3N$

乙：$8\times 0.2N + 9\times 0.5N + 10\times 0.3N = 9.1N$

平均起来甲每枪射中 9.3 环，乙射中 9.1 环，所以甲射手的技术要好些.

因此，这里的计算公式为

$$m = 8\times p_8 + 9\times p_9 + 10\times p_{10}$$

这是对击中环数的加权平均，其权正好取为相应的概率.

二、离散型场合

受上面的问题启发，对一般离散型随机变量，我们可引进如下定义.

定义 4.1.1 设 ξ 为一离散型随机变量，它取值 x_1，x_2，x_3，…对应的概率为 p_1，p_2，p_3，…如果级数

$$\sum_{i=1}^{\infty} x_i p_i \tag{4.1.3}$$

绝对收敛，则把它称为 ξ 的**数学期望**(mathematical expectation)，简称**期望**、**期望值**或**均值** (mean)，记作 $E\xi$.

当 $\sum_{i=1}^{\infty}|x_i|p_i$ 发散时，则说 ξ 的数学期望不存在.

定义中对级数要求绝对收敛是为了数学处理的方便. 从直观上来讲，它也是合理的：因为诸 x_i 的顺序对随机变量并不是本质的，因而在数学期望的定义中就应允许任意改变 x_i 的次序而不影响其收敛性及其和值，这在数学上就相当于要求级数 (4.1.3) 绝对收敛.

显然数学期望由概率分布唯一确定，以后我们也称为某概率分布的数学期望. 下面来计算一些重要的离散型分布的期望值.

［例 1］ 伯努利分布 事件 A 发生的概率为 p，若以 $\mathbf{1}_A$ 记其示性函数，即 A 发生时取值 1，否则取值 0，则

$$\begin{aligned} E\mathbf{1}_A &= 1\times p + 0\times(1-p) \\ &= p = P(A) \end{aligned} \tag{4.1.4}$$

因此概率 $P(A)$是随机变量 $\mathbf{1}_A$ 的数学期望. 从这个角度看，概率是数学期望的特例.

［例 2］ 二项分布 $p_k=\binom{n}{k}p^k q^{n-k}$，$k=0$，1，2，…，$n$

$$\sum_{k=0}^{n} k p_k = \sum_{k=1}^{n} k\binom{n}{k} p^k q^{n-k}$$

$$= np\sum_{k=1}^{n}\binom{n-1}{k-1}p^{k-1}q^{n-k}$$

$$= np(p+q)^{n-1} = np \tag{4.1.5}$$

［例 3］ 泊松分布 $p_k=\frac{\lambda^k}{k!}e^{-\lambda}$，$k=0$，1，2，…

$$\sum_{k=0}^{\infty}kp_k = \sum_{k=1}^{\infty}k\cdot\frac{\lambda^k}{k!}e^{-\lambda} = \lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}$$

$$= \lambda e^{-\lambda}\cdot e^{\lambda} = \lambda \tag{4.1.6}$$

由此看出，泊松分布的参数 λ 就是它的期望值.

［例 4］ 几何分布 $p_k=q^{k-1}p$，$k=1$，2，…

$$\sum_{k=1}^{\infty}kp_k = \sum_{k=1}^{\infty}kq^{k-1}p = p(1+2q+3q^2+\cdots)$$

$$= p(q+q^2+q^3+\cdots)' = p\left(\frac{q}{1-q}\right)'$$

$$= p\frac{1}{(1-q)^2} = \frac{1}{p} \tag{4.1.7}$$

［例 5］ 随机变量 ξ 取值 $x_k=(-1)^k\frac{2^k}{k}$，$k=1$，2，…对应的概率为 $p_k=\frac{1}{2^k}$，则由于 $p_k\geqslant 0$，$\sum\limits_{k=1}^{\infty}p_k=1$，因此它是概率分布，而且

$$\sum_{k=1}^{\infty}x_kp_k = \sum_{k=1}^{\infty}(-1)^k\frac{1}{k} = -\ln 2$$

但由于

$$\sum_{k=1}^{\infty}|x_k|p_k = \sum_{k=1}^{\infty}\frac{1}{k} = \infty$$

因此 ξ 的数学期望不存在.

从上面例子看到，几种重要的离散型分布，其参数都可由数学期望算得，因此数学期望是一个重要的概念.

三、应用实例

从上面讨论中可以看出，数学期望刻画了随机变量取值的某种平均，有明显的直观含意．在许多问题中，数学期望的概念甚至比概率和随机变量等概念更易被人理解和接受．在概率论发展的早期，人们更常用的概念和工具是数学期望．下面的例子将有助读者对这个重要概念的深入理解．

［例 6］（押宝）押宝是赌博的一种，它以种种形式在世界各地流行，吞夺大量财富，现举一例以揭穿其本质．

在我国南方流行一种称为“捉水鸡”的押宝，其规则如下：由庄家摸出一只棋子，放在密闭的盒中，这只棋子可以是红的或黑的将、士、象、车、马、炮之一．赌客们把钱押在一块写有上述十二个字（六个红字、六个黑字）的台面的某个字上．押定后，庄家揭开盒子露出原来的棋子．凡押中者（字和颜色都对）以 1 比 10 得到赏金，不中者其押金归庄家．

为对这种押宝的实质有个了解，最好考察一个赌徒当他押上 1 元赌注之后的期望所得．显然其分布列为

$$\begin{pmatrix} 0 & 11 \\ \frac{11}{12} & \frac{1}{12} \end{pmatrix} \tag{4.1.8}$$

因此其数学期望为$\frac{11}{12}$.

由于支出（1 元）和期望收入$\left(\frac{11}{12}\text{元}\right)$不等，因此这是不公平的赌博，它明显对庄家有利．数学期望的概念帮助我们认请了这点．

事实上，当一个赌徒走进一座赌场时，他面临的都是这种不公平的赌博，否则赌场的巨额开销和业主的高额利润从何而来呢？

普及概率知识有助于杜绝赌博现象．

［例 7］（彩票）彩票的发行，数额巨大，其实质如何呢？请

看一则实例：发行彩票10万张，每张1元．设头等奖1个，奖金1万元；二等奖2个，奖金各5仟元；三等奖10个，奖金各1仟元；四等奖100个，奖金各1佰元；五等奖1000个，奖金各10元．

还是算算每张彩票的期望所得．这时分布列为

$$\begin{pmatrix} 10000 & 5000 & 1000 & 100 & 10 & 0 \\ \frac{1}{10^5} & \frac{2}{10^5} & \frac{10}{10^5} & \frac{100}{10^5} & \frac{1000}{10^5} & * \end{pmatrix} \quad (4.1.9)$$

不难算出其期望值为0.5元，即大约能收回一半．

在我国，类似的发行只有当收益主要用于公益事业时才允许．另外，有奖储蓄在保证储户基本收入的前提下，用部分利息摇奖，其分析方法类同．

[例8]　（保险）这里主要问题是如何确定保险费．在保险学中，收取保险费的原则是：被保险人交的“纯保险费”与他们所能得到的赔偿金的期望值相等．

因此，若出事的概率为 p，有 N 个人参加保险，则每人交的纯保险费 a 与出事赔偿金 b，应有下面关系：

$$Na = \sum_{k=0}^{N} p^k(1-p)^{N-k} \cdot kb$$

即 $a=pb$，正如所预期的．

保险的种类繁多，但上述原则不变，因此保险费的计算离不开概率论．

[例9]　（投资之决策）投资总具有一定风险，因此在选择投资方向时，计算其期望收益常是可供考虑的决策方法之一．下面是一个大为简化的例子．

某人有10万元现金，想投资于某项目，预估成功的机会为30%，可得利润8万元，失败的机会为70%，将损失2万元．若存入银行，同期间的利率为5%，问是否应作此项投资？

以 ξ 记投资利润，则

$$E\xi = 8 \times 0.3 - 2 \times 0.7 = 1(\text{万元})$$

而存入银行的利息为 $10\times5\%=0.5$（万元），因此从期望收益的角度看，应选择投资，当然这时要冒一定的风险.

［例 10］（一种验血新技术）在一个人数很多的单位中普查某种疾病，N 个人去验血. 对这些人的血的化验可以用两种办法进行.（1）每个人的血分别化验，这时需要化验 N 次；（2）把 k 个人的血混在一起进行化验. 如果结果是阴性的，那么对这 k 个人只作一次检验就够了；如果结果是阳性的，那么必须对这 k 个人再逐个分别化验，这时对这 k 个人共需作 $k+1$ 次化验. 假定对所有的人来说，化验是阳性反应的概率都是 p，而且这些人的反应是独立的. 我们来说明在 p 相当小的场合，采用办法（2）能减少化验的次数.

若记 $q=1-p$，则 k 个人的混血呈阳性反应的概率为 $1-q^k$，用（2）的方法验血时，每个人的血需要化验的次数 ξ 是随机变量，其分布列为

$$\begin{pmatrix} \frac{1}{k} & 1+\frac{1}{k} \\ q^k & 1-q^k \end{pmatrix}$$

因此

$$E\xi=\frac{1}{k}\cdot q^k+\left(1+\frac{1}{k}\right)(1-q^k)=1-q^k+\frac{1}{k}$$

N 个人需要的化验次数的期望值为 $N\left(1-q^k+\frac{1}{k}\right)$，当 $q^k-\frac{1}{k}>0$ 时，就能减少验血次数. 例如当 $p=0.1$ 时，取 $k=4$，则 $q^k-\frac{1}{k}=0.4$. 用（2）法平均能减少 40%的工作量. 显然 p 愈小，用这种方法愈有利. 当 p 已知时，还可以选定整数 k_0，使 $E\xi$ 达到最小，把 k_0 个人分为一组就最能节省化验次数.

四、连续型场合

下面我们转入考虑连续型随机变量的数学期望. 设随机变量

ξ 有密度函数 $p(x)$，取很密的分点 $x_0<x_1<x_2<\cdots<x_n$，则 ξ 落在 $[x_i, x_{i+1})$ 中的概率近似地等于 $p(x_i)(x_{i+1}-x_i)$，因此 ξ 与以概率 $p(x_i)(x_{i+1}-x_i)$ 取值 x_i 的离散型随机变量近似，而这离散型随机变量的数学期望为

$$\sum_i x_i p(x_i)(x_{i+1}-x_i)$$

上式是积分 $\int_{-\infty}^{\infty} xp(x)\mathrm{d}x$ 的渐近和式，这个直观的考虑启发我们引进如下定义：

定义 4.1.2 设 ξ 为具有密度函数 $p(x)$ 的连续型随机变量，当积分 $\int_{-\infty}^{\infty} xp(x)\mathrm{d}x$ 绝对收敛时，我们称它为 ξ 的**数学期望**（或**均值**），记作 $E\xi$，即

$$E\xi=\int_{-\infty}^{\infty} xp(x)\mathrm{d}x \tag{4.1.10}$$

显然这里定义的数学期望也只与分布有关．下面计算一些重要的连续型分布的数学期望．

［例 11］ 正态分布 $N(a, \sigma^2)$.

$$\begin{aligned}\int_{-\infty}^{\infty} xp(x)\mathrm{d}x &= \int_{-\infty}^{\infty} x\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-(x-a)^2/(2\sigma^2)}\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}(\sigma z+a)\mathrm{e}^{-z^2/2}\mathrm{d}z \\ &= \frac{a}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-z^2/2}\mathrm{d}z = a\end{aligned} \tag{4.1.11}$$

可见 $N(a, \sigma^2)$ 中的参数 a 正是它的数学期望．

［例 12］ 指数分布 $p(x)=\lambda\mathrm{e}^{-\lambda x}$，$x\geqslant 0$.

$$\int_0^{\infty} x\lambda\mathrm{e}^{-\lambda x}\mathrm{d}x = -\int_0^{\infty} x\mathrm{d}\mathrm{e}^{-\lambda x} = \int_0^{\infty}\mathrm{e}^{-\lambda x}\mathrm{d}x = \frac{1}{\lambda} \tag{4.1.12}$$

［例 13］ 柯西分布 $p(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$.

由于

$$\int_{-\infty}^{\infty}|x|\cdot\frac{1}{\pi(1+x^2)}\mathrm{d}x=\infty$$

因此柯西分布的数学期望不存在.

五、一般场合

我们已经对离散型随机变量及连续型随机变量分别定义了数学期望，现在自然希望找到一种能适合一切随机变量的数学期望定义，并把上述两种情况作为特例. 为了做到这点，需要利用斯蒂尔切斯（Stieltjes）积分①.

若随机变量 ξ 的分布函数为 $F(x)$，类似于连续型随机变量的场合，作很密的分割 $x_0<x_1<x_2<\cdots<x_n$，则 ξ 落在 $[x_i,x_{i+1})$ 中的概率等于 $F(x_{i+1})-F(x_i)$，因此 ξ 与以概率 $F(x_{i+1})-F(x_i)$ 取值 x_i 的离散型随机变量近似，而后者的数学期望为

$$\sum_i x_i[F(x_{i+1})-F(x_i)]$$

注意到上式是斯蒂尔切斯积分 $\int_{-\infty}^{\infty}x\mathrm{d}F(x)$ 的渐近和式，这启发我们引进如下定义：

定义 4.1.3 若 ξ 的分布函数为 $F(x)$，则定义

$$E\xi=\int_{-\infty}^{\infty}x\mathrm{d}F(x) \tag{4.1.13}$$

为 ξ 的**数学期望**. 这里我们还是要求上述积分绝对收敛，否则数学期望不存在.

关于斯蒂尔切斯积分

$$I=\int_{-\infty}^{\infty}g(x)\mathrm{d}F(x) \tag{4.1.14}$$

我们仅列举它的如下性质：

---

① 如果读者不熟悉斯蒂尔切斯积分而又希望对它有所了解，可参看复旦大学数学系，实变数函数论与泛函分析概要（第二版），上海科技出版社，1963，第四章. 不过，为读懂本书其余部分，只要承认后面所述的少数事实就可以了.

(i) 当 $F(x)$为跳跃函数，在 x_i $(i=1, 2, \cdots)$ 具有跃度 p_i 时，上面积分化为无穷级数

$$I = \sum_i g(x_i) p_i$$

(ii) 当 $F(x)$存在导数 $F'(x)=p(x)$时，积分(4.1.14)化为普通积分

$$I = \int_{-\infty}^{\infty} g(x)p(x)\mathrm{d}x$$

(iii) 线性性质

$$\int_{-\infty}^{\infty} [ag_1(x) + bg_2(x)]\mathrm{d}F(x)$$

$$= a\int_{-\infty}^{\infty} g_1(x)\mathrm{d}F(x) + b\int_{-\infty}^{\infty} g_2(x)\mathrm{d}F(x)$$

(iv)

$$\int_{-\infty}^{\infty} g(x)\mathrm{d}[aF_1(x) + bF_2(x)]$$

$$= a\int_{-\infty}^{\infty} g(x)\mathrm{d}F_1(x) + b\int_{-\infty}^{\infty} g(x)\mathrm{d}F_2(x)$$

(v) $$\int_a^b g(x)\mathrm{d}F(x) = \int_a^c g(x)\mathrm{d}F(x) + \int_c^b g(x)\mathrm{d}F(x)$$

$(a \leqslant c \leqslant b)$

(vi) 若 $g(x) \geqslant 0$，$F(x)$单调不减，$b>a$，则

$$\int_a^b g(x)\mathrm{d}F(x) \geqslant 0$$

从前两个性质，我们知道定义式（4.1.13）的确能包含（4.1.3）及（4.1.10）作为特例.

六、随机变量函数的数学期望

下面讨论随机变量的函数 $\eta=g(\xi)$ 的数学期望的定义，这里 $g(x)$是一元博雷尔函数，从上章 § 3 的讨论已知 η 是随机变量. 通过类似于引进（4.1.13）的推理，似应定义 $g(\xi)$ 的数学期望为

$$Eg(\xi) = \int_{-\infty}^{\infty} g(x)\mathrm{d}F_{\xi}(x) \qquad (4.1.15)$$

这里当然还是要求这个积分绝对收敛.

但是，另一方面，因为η是随机变量，也有分布函数，记之为$F_{\eta}(x)$，则按一般随机变量数学期望的定义式（4.1.13）又应有

$$E\eta = \int_{-\infty}^{\infty} y\mathrm{d}F_{\eta}(y) \qquad (4.1.16)$$

因此，这两个积分应该相等. 事实上，这两个积分的确相等，我们把这个事实写成定理的形式.

定理 4.1.1 若$g(x)$是一元博雷尔函数，而$\eta=g(\xi)$，则

$$\int_{-\infty}^{\infty} y\mathrm{d}F_{\eta}(y) = \int_{-\infty}^{\infty} g(x)\mathrm{d}F_{\xi}(x) \qquad (4.1.17)$$

即这两个积分中，若有一个存在，则另一个也存在，而且两者相等.

这个定理的证明要用到积分论，超出了本课程范围. 我们只能列出结论，并指出它的重要性：一方面，它消除了随机变量的数学期望定义中所出现的表面矛盾；另一方面在计算随机变量函数的数学期望时也带来很大的方便，我们无须先计算η的分布函数$F_{\eta}(y)$再求其数学期望，而可以直接从ξ的分布函数$F_{\xi}(x)$出发利用（4.1.15）来计算.

公式（4.1.15）有十分明显的直观解释. 在离散型场合，公式（4.1.15）化为

$$Eg(\xi) = \sum_{i=1}^{\infty} g(x_i)p(x_i) \qquad (4.1.18)$$

我们以一个浅显的例子来说明它的含意.

某零售商店出售某种小商品，每销售一件可赚 1.5 元，关心的是它的利润η. 由于利润与销售量ξ成正比，因此要知道销售量. 比较合理的做法是假定ξ为随机变量，服从一定分布. 为简单起见，假定每天卖出 0，1，2，3 件的概率分别为 0.4，0.3，0.2，0.1. 这时希望计算平均利润$E\eta$.

第一种计算方法是把η看作随机变量，求出其概率分布，再按一般数学期望公式计算．这时η的分布列为

$$\begin{pmatrix} 0 & 1.5 & 3 & 4.5 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{pmatrix}$$

因此

$$\begin{aligned} E\eta &= 0 \times 0.4 + 1.5 \times 0.3 + 3 \times 0.2 + 4.5 \times 0.1 \\ &= 1.5(\text{元}) \end{aligned}$$

这种算法就是按公式（4.1.16）的思路进行的.

另一种算法把η看作ξ的函数$\eta=1.5\xi$，由于ξ的分布列早就知道为

$$\begin{pmatrix} 0 & 1 & 2 & 3 \\ 0.4 & 0.3 & 0.2 & 0.1 \end{pmatrix}$$

因此按（4.1.18）式（也即按（4.1.15）式）

$$\begin{aligned} E\eta = E(1.5\xi) &= 1.5 \times 0 \times 0.4 + 1.5 \times 1 \times 0.3 + 1.5 \times 2 \\ &\quad \times 0.2 + 1.5 \times 3 \times 0.1 \\ &= 1.5(\text{元}) \end{aligned}$$

正如预期的，两个答案相同，这也为定理 4.1.1 提供了例证.

下面的例子更接近于实际，不过其方法实质是一样的.

［例 14］（报童问题）设某报童每日的潜在卖报数ζ服从参数为λ的泊松分布. 如果每卖出一份报可得报酬a，卖不掉而退回则每份赔偿b. 若某日该报童买进n份报，试求其期望所得. 进一步，还要求最佳的卖出份数n.

［解］　若记其真正卖报数为ξ，则ξ与ζ的关系如下：

$$\xi = \begin{cases} \zeta, & \zeta < n \\ n, & \zeta \geqslant n \end{cases}$$

这里的ξ服从截尾泊松分布，即

$$P\{\xi = k\} = \begin{cases} \dfrac{\lambda^k}{k!}e^{-\lambda}, & k < n \\ \displaystyle\sum_{i=n}^{\infty} \dfrac{\lambda^i}{i!}e^{-\lambda}, & k = n \end{cases}$$

记所得为 η，则 η 与 ξ 的关系如下：

$$\eta = g(\xi) = \begin{cases} an, & \xi = n \\ a\xi - b(n-\xi), & \xi < n \end{cases}$$

因此由（4.1.18），期望所得为

$$M(n) = Eg(\xi) = \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} e^{-\lambda} [ka - (n-k)b] + \left(\sum_{k=n}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} \right) na$$

$$= (a+b)\lambda \sum_{k=0}^{n-2} \frac{\lambda^k}{k!} e^{-\lambda} - n(a+b) \sum_{k=0}^{n-1} \frac{\lambda^k}{k!} e^{-\lambda} + na$$

这个问题的最终解决是当 a，b，λ 给定后，求 n 使 $M(n)$ 达到极大，这是一个典型的最优化问题，也是报童问题的正确解.

一般计算泊松分布的部分和可用下列公式：

$$\sum_{k=r}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = \int_0^1 \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx = \frac{1}{\Gamma(r)} \int_0^{\lambda} y^{r-1} e^{-y} dy \tag{4.1.19}$$

其数值可在数学表中查到. 公式（4.1.19）是埃尔兰分布的一个应用，也可直接证明。

另一重要的特殊情况是 ξ 具有密度函数 $p(x)$ 的场合，此时

$$E\eta = Eg(\xi) = \int_{-\infty}^{\infty} g(x)p(x) dx \tag{4.1.20}$$

这个结论经常用到. 下面是一个例子，这个例子也顺带说明可以利用随机变量的期望值来作出某种最优决策.

［例 15］ 假定在国际市场上每年对我国某种出口商品的需求量是随机变量 ξ（吨），它服从［2000，4000］上均匀分布. 设每售出这种商品 1 吨，可为国家挣得外汇 3 万元，但假如销售不出而屯积于仓库，则每吨需浪费保养费 1 万元，问题是要确定应组织多少货源，才能使国家的收益最大？

若以 y 记预备某年出口的此种商品量（显然可以只考虑 $2000 \leqslant y \leqslant 4000$ 的情况），则收益（单位：万元）

$$\eta = H(\xi) = \begin{cases} 3y, & \text{当 } \xi \geqslant y \text{ 时} \\ 3\xi - (y-\xi), & \text{当 } \xi < y \text{ 时} \end{cases}$$

为了求得 $E\eta$，利用公式（4.1.20），

$$E\eta=\int_{-\infty}^{\infty}H(x)p(x)\mathrm{d}x=\frac{1}{2000}\int_{2000}^{4000}H(x)\mathrm{d}x$$
$$=\frac{1}{2000}\int_{2000}^{y}(4x-y)\mathrm{d}x+\frac{1}{2000}\int_{y}^{4000}3y\mathrm{d}x$$
$$=\frac{1}{1000}(-y^2+7000y-4\times10^6)$$

此式当 $y=3500$ 时达到最大，因此组织3500吨此种商品是最好的决策.

七、多维场合

可以把上面结果推广到随机向量的场合，若 $(\xi_1,\cdots,\xi_n)$ 的分布函数为 $F(x_1,\cdots,x_n)$，而 $g(x_1,\cdots,x_n)$ 为 n 元博雷尔函数，则

$$Eg(\xi_1,\cdots,\xi_n)=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}g(x_1,\cdots,x_n)\mathrm{d}F(x_1,\cdots,x_n)$$

特别地，

$$E\xi_1=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_1\mathrm{d}F(x_1,\cdots,x_n)=\int_{-\infty}^{\infty}x_1\mathrm{d}F_1(x_1)$$

其中 $F_1(x_1)$ 是 ξ_1 的分布函数. 一般地，引进如下定义

定义 4.1.4 随机向量 $(\xi_1,\xi_2,\cdots,\xi_n)$ 的**数学期望**为 $(E\xi_1,E\xi_2,\cdots,E\xi_n)$，其中

$$E\xi_i=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}x_i\mathrm{d}F(x_1,\cdots,x_n)=\int_{-\infty}^{\infty}x_i\mathrm{d}F_i(x_i)$$

这里 $F_i(x_i)$ 是 ξ_i 的分布函数 $(i=1,2,\cdots,n)$.

因此二元正态分布（3.2.22）中的参数 (a_1,a_2) 正是随机向量的数学期望.

八、数学期望的基本性质

性质 1 若 $a\leqslant\xi\leqslant b$，则 $a\leqslant E\xi\leqslant b$. 特别地 $Ec=c$，这里 a,b,c 是常数.

性质 2 线性性质：对任意常数 c_i，$i=1,\cdots,n$ 及 b，有

$$E(\sum_{i=1}^{n}c_i\xi_i + b) = \sum_{i=1}^{n}c_iE\xi_i + b$$

这两个性质的证明是明显的．性质 2 对数学期望的计算很有好处．

［例 16］　求超几何分布

$$p_m = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}, \quad m = 0,1,\cdots,n$$

的数学期望．

当然可以用$\sum_{k=0}^{n}kp_k$直接求出，但也可用下面方法来计算．

设想一个相应的不放回抽样，令

$$\xi_i = \begin{cases} 1, & \text{第 } i \text{ 次抽得废品} \\ 0, & \text{第 } i \text{ 次抽得好品} \end{cases}$$

则 $P\{\xi_i = 1\} = \frac{M}{N}$，因此 $E\xi_i = \frac{M}{N}$，而 $\xi = \xi_1 + \cdots + \xi_n$ 表示 n 次抽样中抽出的废品数，它服从超几何分布，利用性质 2 得到

$$E\xi = E\xi_1 + \cdots + E\xi_n = \frac{nM}{N}$$

§2. 方差，相关系数，矩

一、方差

数学期望是随机变量的一个重要数字特征，它表示了随机变量取值的平均水平，从一个角度描述了随机变量．但是从下面例子马上可以看出，单用数学期望描述随机变量通常是不够的．

再考察两个射手，他们的射击技术用下表表出：

丙射手

击中环数	8	9	10
概　　率	0.1	0.8	0.1

丁射手

击中环数	8	9	10
概　　率	0.4	0.2	0.4

显然，他们每射一枪的期望值都是 9 环. 不过，他们之间的技术差异是明显的，有必要进一步刻画.

细察之下发现，丙射手的射击大部分集中在均值 9 环，而丁射手则散布度比较大.

一般地，当我们考察一组观察值 x_1，x_2，…，x_n 时，其平均值 $\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$ 给予我们一个关于这组数据中心的表征. 接下来我们注意的便是这组数据对于平均值的偏离，也就是 $x_1-\bar{x}$，$x_2-\bar{x}$，…，$x_n-\bar{x}$. 当然我们也想用一个数字来刻画这种偏离. 首先想到的可能是把这些偏离值求和，但是马上发现这是没有意义的，因为 $\sum_{i=1}^{n}(x_i-\bar{x})\equiv 0$，这是平均值的性质之一，它来自正负偏离相互抵消. 这时可供选择的指标有两个，一个便是平均绝对偏差

$$AD=\frac{1}{n}\sum_{i=1}^{n}|x_i-\bar{x}|$$

可惜，这个量从数学上不太容易处理，因此最后还是让位于另一个指标——方差

$$\sigma^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

上述讨论容易过渡到带有“权”的场合，从而到离散型随机变量的场合，即对于数学期望的偏离的平方的加权平均

$$\sum_{i=1}^{n}(x_i-EX)^2p_i$$

这些直观讨论启发我们引进如下定义：

定义 4.2.1 若 $E(\xi-E\xi)^2$ 存在，则称它为随机变量 ξ 的**方差** (variance)，并记为 $D\xi$，而 $\sqrt{D\xi}$ 称为**根方差**、**均方差**或更多地称为**标准差** (standard deviation).

利用数学期望的线性性质，

$$\begin{aligned}D\xi&=E(\xi-E\xi)^2=E[\xi^2-2\xi\cdot E\xi+(E\xi)^2]\\&=E\xi^2-2E\xi\cdot E\xi+(E\xi)^2=E\xi^2-(E\xi)^2\end{aligned}\qquad(4.2.1)$$

方差描述了随机变量对于其数学期望的离散程度，在概率论和数理统计中十分重要．例如本节开头讨论的丙、丁两射手，丁射手的方差大，即技术不够“稳定”；而丙射手方差小，较“稳定”．这与习惯看法是一致的．

当然方差也由概率分布完全确定．下面计算一些重要分布的方差，为书写方便，一律假定相应的随机变量为 ξ.

［例 1］ 伯努利分布

$$E\xi^2 = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

$$D\xi = E\xi^2 - (E\xi)^2 = p - p^2 = pq \tag{4.2.2}$$

［例 2］ 二项分布

$$E\xi^2 = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k q^{n-k} = npq + n^2 p^2$$

$$D\xi = E\xi^2 - (E\xi)^2 = npq + n^2 p^2 - n^2 p^2 = npq \tag{4.2.3}$$

［例 3］ 泊松分布

$$E\xi^2 = \sum_{k=0}^{\infty} k^2 p_k = \sum_{k=1}^{\infty} k^2 \cdot \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} k \frac{\lambda^k}{(k-1)!} e^{-\lambda}$$

$$= \lambda \sum_{k=0}^{\infty} (k+1) \frac{\lambda^k}{k!} e^{-\lambda} = \lambda^2 + \lambda$$

$$D\xi = E\xi^2 - (E\xi)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda \tag{4.2.4}$$

［例 4］ 均匀分布 U $[a, b]$

$$E\xi = \int_a^b x \frac{1}{b-a} dx = \frac{b+a}{2} \tag{4.2.5}$$

$$E\xi^2 = \int_a^b x^2 \frac{1}{b-a} dx = \frac{b^2 + ab + a^2}{3}$$

$$D\xi = E\xi^2 - (E\xi)^2 = \frac{b^2 + ab + a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{(b-a)^2}{12} \tag{4.2.6}$$

［例 5］ 正态分布 $N(a, \sigma^2)$

$$D\xi = \int_{-\infty}^{\infty} (x-a)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-a)^2/(2\sigma^2)} dx$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} z^2 e^{-z^2/2} dz$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\left[\left(-ze^{-z^2/2}\right)\Big|_{-\infty}^{\infty}+\int_{-\infty}^{\infty} e^{-z^2/2} dz\right]$$

$$=\frac{\sigma^2}{\sqrt{2\pi}}\sqrt{2\pi}=\sigma^2 \tag{4.2.7}$$

这样，我们阐明了正态分布中第二个参数 σ 的概率意义，它就是标准差；而正态分布也由它的数学期望及标准差唯一确定.

下面讨论方差的性质

性质 1 常数的方差为 0.

性质 2 $D(\xi+c)=D\xi$，这里 c 是常数.

性质 3 $D(c\xi)=c^2D\xi$，这里 c 是常数.

对于随机变量 ξ，若它的数学期望 $E\xi$ 及方差 $D\xi$ 都存在，而且 $D\xi>0$，有时要考虑**标准化**了的随机变量

$$\xi^*=\frac{\xi-E\xi}{\sqrt{D\xi}}$$

显然 $E\xi^*=0$，$D\xi^*=1$，这正是称 ξ^* 为标准化随机变量的理由.

性质 4 若 $c\neq E\xi$，则 $D\xi<E(\xi-c)^2$ (4.2.8)

[证明] 因为

$$D\xi=E(\xi-E\xi)^2=E(\xi-c)^2-(c-E\xi)^2 \tag{4.2.9}$$

这个性质表明数学期望具有一个重要的极值性质：在 $E(\xi-c)^2$ 中，当 $c=E\xi$ 时达到极小；这也说明在 $D\xi$ 的定义中取 $c=E\xi$ 的合理性.

二、车贝晓夫（Чебышев）不等式

概率论中有许多不等式，下面的车贝晓夫不等式是其中最基本和最重要的一个.

车贝晓夫不等式 对于任何具有有限方差的随机变量 ξ，都有

$$P\{|\xi-E\xi|\geqslant\varepsilon\}\leqslant\frac{D\xi}{\varepsilon^2} \tag{4.2.10}$$

其中 ε 是任一正数.

［证明］ 若 $F(x)$是 ξ 的分布函数,则显然有

$$\begin{aligned} D\xi &=\int_{-\infty}^{\infty}(x-E\xi)^2\mathrm{d}F(x) \\ &\geqslant \int_{|x-E\xi|\geqslant\varepsilon}(x-E\xi)^2\mathrm{d}F(x)\geqslant \int_{|x-E\xi|\geqslant\varepsilon}\varepsilon^2\mathrm{d}F(x) \\ &=\varepsilon^2 P\ \{|\xi-E\xi|\geqslant\varepsilon\} \end{aligned} \tag{4.2.11}$$

这就证得了不等式（4.2.10）. 有时把（4.2.10）改写成

$$P\{|\xi-E\xi|<\varepsilon\}\geqslant 1-\frac{D\xi}{\varepsilon^2} \tag{4.2.12}$$

或

$$P\left\{\left|\frac{\xi-E\xi}{\sqrt{D\xi}}\right|\geqslant\delta\right\}\leqslant\frac{1}{\delta^2} \tag{4.2.13}$$

车贝晓夫不等式利用随机变量 ξ 的数学期望 $E\xi$ 及方差 $D\xi=\sigma^2$ 对 ξ 的概率分布进行估计. 例如(4.2.13)断言不管 ξ 的分布是什么，ξ 落在（$E\xi-\sigma\delta$, $E\xi+\sigma\delta$）中的概率均不小于 $1-\frac{1}{\delta^2}$. 因为车贝晓夫不等式只利用数学期望及方差就描述了随机变量的变化情况，因此它在理论研究及实际应用中都很有价值.

从车贝晓夫不等式还可以看出，当方差愈小时，事件 $\{|\xi-E\xi|\geqslant\varepsilon\}$的概率也愈小，从这里可以看出方差是描述随机变量与其期望值离散程度的一个量,这与我们以前的理解完全一致.

特别地，若 $D\xi=0$，则对任意 $\varepsilon>0$，恒有$P\{|\xi-E\xi|\geqslant\varepsilon\}=0$，因此 $P\{\xi\neq E\xi\}=0$，即 $P\{\xi=E\xi\}=1$，所以方差为零的随机变量是常数.

三、相关系数

对于随机向量 $\xi=(\xi_1, \xi_2, \cdots, \xi_n)$，定义它的**方差**为（$D\xi_1$,

$D\xi_2$, …, $D\xi_n$).

方差反映了随机向量各个分量对于各自的数学期望的离散程度，它对于了解随机向量的分布有一定帮助. 但对于随机向量，我们除了关心它的每个分量的情况外，还希望知道各个分量之间的联系，这光靠数学期望与方差是办不到的. 下面引进的量则能起这个作用.

定义 4.2.2 称

$$b_{ij}=\operatorname{cov}(\xi_i,\xi_j)=E[(\xi_i-E\xi_i)(\xi_j-E\xi_j)]$$
$$i,j=1,2,\cdots,n \tag{4.2.14}$$

为 ξ_i 与 ξ_j 的协方差 (covariance).

不难验算

$$\operatorname{cov}(\xi_i,\xi_j)=E\xi_i\xi_j-E\xi_i\cdot E\xi_j \tag{4.2.15}$$

$$D\Big(\sum_{i=1}^{n}\xi_i\Big)=\sum_{i=1}^{n}D\xi_i+2\sum_{1\leqslant i<j\leqslant n}\operatorname{cov}(\xi_i,\xi_j) \tag{4.2.16}$$

特别地

$$D(\xi+\eta)=D\xi+D\eta+2\operatorname{cov}(\xi,\eta) \tag{4.2.17}$$

方差是协方差的特例，显然 $b_{ii}=D\xi_i$，矩阵

$$\boldsymbol{B}=\begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{pmatrix} \tag{4.2.18}$$

称为 ξ 的**协方差矩阵**，简记作 $D\boldsymbol{\xi}$，显然这是一个对称矩阵.

此外，对任何实数 $t_j(j=1,2,\cdots,n)$有

$$\int\cdots\int\Big\{\sum_{j=1}^{n}t_j(x_j-E\xi_j)\Big\}^2\mathrm{d}F(x_1,\cdots,x_n)=\sum_{j,k}b_{jk}t_jt_k\geqslant 0$$

因此 $\boldsymbol{B}$ 是一个非负定矩阵，所以若以 $\det\boldsymbol{B}$ 记 $\boldsymbol{B}$ 的行列式，则有

$$\det\boldsymbol{B}\geqslant 0$$

更常用的是如下“标准化”了的协方差.

定义 4.2.3 称

$$r_{ij}=\frac{\mathrm{cov}\ (\xi_i,\ \xi_j)}{\sqrt{D\xi_i}\sqrt{D\xi_j}} \tag{4.2.19}$$

为 ξ_i 与 ξ_j 的**相关系数**(correlation coefficient),这里当然要求 $D\xi_i$ 与 $D\xi_j$ 不为零. 相关系数也就是标准化的随机变量 $\frac{\xi_i-E\xi_i}{\sqrt{D\xi_i}}$ 与 $\frac{\xi_j-E\xi_j}{\sqrt{D\xi_j}}$ 的协方差.

以后定义常数与任何随机变量的相关系数为 0.

[例 6] 多项分布(3.2.6)

显然,$\xi_i\sim B(n,p_i),i=1,2,\cdots,r$. 因此

$$E\xi_i=np_i,D\xi_i=np_i(1-p_i)$$

为求协方差或相关系数,可用下面技巧:注意到

$$\xi_i+\xi_j\sim B(n,p_i+p_j)$$

因此

$$E(\xi_i+\xi_j)=n(p_i+p_j),\quad D(\xi_i+\xi_j)=n(p_i+p_j)(1-p_i-p_j)$$

由于

$$\begin{aligned}D(\xi_i+\xi_j)&=D\xi_i+D\xi_j+2\mathrm{cov}(\xi_i,\xi_j)\\&=np_i(1-p_i)+np_j(1-p_j)+2\mathrm{cov}(\xi_i,\xi_j)\end{aligned}$$

可以得到

$$\mathrm{cov}(\xi_i,\xi_j)=-np_ip_j \tag{4.2.20}$$

相关系数为

$$\begin{aligned}r_{ij}&=\frac{-p_ip_j}{\sqrt{p_ip_j(1-p_i)(1-p_j)}}\\&=-\sqrt{\frac{p_ip_j}{(1-p_i)(1-p_j)}}\end{aligned}$$

现在讨论相关系数的性质. 因为相关系数只与两个随机变量

有关，下面讨论将假定这两个随机变量是ξ与η，它们的相关系数记为r.

先证明一条常用的定理.

定理 4.2.1（柯西—施瓦兹不等式） 对任意随机变量ξ与η都有

$$|E\xi\eta|^2\leqslant E\xi^2\cdot E\eta^2 \tag{4.2.21}$$

等式成立当且仅当

$$P\{\eta=t_0\xi\}=1 \tag{4.2.22}$$

这里t_0是某一个常数.

［证明］ 对任意实数t，定义$u(t)=E(t\xi-\eta)^2=t^2E\xi^2-2tE\xi\eta+E\eta^2$，显然对一切$t$，$u(t)\geqslant 0$，因此二次方程$u(t)=0$或者没有实根或者有一个重根. 所以

$$[E\xi\eta]^2-E\xi^2\cdot E\eta^2\leqslant 0$$

这正是(4.2.21). 此外，方程$u(t)=0$有一个重根t_0存在的充要条件是

$$[E\xi\eta]^2-E\xi^2E\eta^2=0$$

这时$E(t_0\xi-\eta)^2=0$，因此

$$P\{t_0\xi-\eta=0\}=1$$

这就是(4.2.22)式. 定理证毕.

由定理 4.2.1 立即可以推出，若两随机变量的方差存在，则它们的协方差也存在.

把定理 4.2.1 应用到随机变量$\dfrac{\xi-E\xi}{\sqrt{D\xi}}$及$\dfrac{\eta-E\eta}{\sqrt{D\eta}}$，可以得到相关系数的如下重要性质：

性质 1 对相关系数r，成立

$$|r|\leqslant 1 \tag{4.2.23}$$

并且$r=1$当且仅当

$$P\left\{\frac{\xi-E\xi}{\sqrt{D\xi}}=\frac{\eta-E\eta}{\sqrt{D\eta}}\right\}=1 \tag{4.2.24}$$

而 $r=-1$ 当且仅当

$$P\left\{\frac{\xi-E\xi}{\sqrt{D\xi}}=-\frac{\eta-E\eta}{\sqrt{D\eta}}\right\}=1 \qquad (4.2.25)$$

性质 1 表明，当 $r=\pm1$ 时，ξ 与 η 存在着线性关系，这时如果给定一个随机变量之值，另一个随机变量的值便完全决定.

有线性关系是一个极端，另一个极端是 $r=0$ 的场合. 为此我们引进：

定义 4.2.4 若随机变量 ξ 与 η 的相关系数 $r=0$，则我们称 ξ 与 η **不相关**.

性质 2 对随机变量 ξ 与 η，下面事实是等价的：

(i) $\mathrm{cov}(\xi,\eta)=0$；

(ii) ξ 与 η 不相关；

(iii) $E\xi\eta=E\xi E\eta$；

(iv) $D(\xi+\eta)=D\xi+D\eta$.

[证明] 显然(i)与(ii)是等价的. 由于

$$\mathrm{cov}(\xi,\eta)=E(\xi-E\xi)(\eta-E\eta)=E\xi\eta-E\xi\cdot E\eta$$

因此(i)与(iii)等价. 又由于

$$D(\xi+\eta)=E(\xi-E\xi+\eta-E\eta)^2=D\xi+D\eta+2\,\mathrm{cov}(\xi,\eta)$$

因此(i)与(iv)等价.

独立性和不相关性都是随机变量间联系“薄弱”的一种反映，自然希望知道这两个概念之间的联系. 首先，我们有

性质 3 若 ξ 与 η 独立，则 ξ 与 η 不相关.

[证明] 我们只对连续型随机变量给出证明.

因为 ξ 与 η 独立，故其密度函数 $p(x,y)=p_1(x)p_2(y)$，因此

$$\mathrm{cov}(\xi,\eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-E\xi)(y-E\eta)p(x,y)\mathrm{d}x\,\mathrm{d}y$$

$$=\int_{-\infty}^{\infty}(x-E\xi)p_1(x)\mathrm{d}x\cdot\int_{-\infty}^{\infty}(y-E\eta)p_2(y)\mathrm{d}y=0$$

结合性质 2 及性质 3，可以得到，若 ξ 与 η 独立，则 $E\xi\eta=E\xi\cdot E\eta$ 及 $D(\xi+\eta)=D\xi+D\eta$ 成立. 同样的论证可以证明这个结论在

n 个随机变量的场合也成立，即若 $\xi_1,\xi_2,\cdots,\xi_n$ 是相互独立的随机变量，则

$$E\xi_1\xi_2\cdots\xi_n=E\xi_1E\xi_2\cdots E\xi_n \tag{4.2.26}$$

$$D(\xi_1+\xi_2+\cdots+\xi_n)=D\xi_1+D\xi_2+\cdots+D\xi_n \tag{4.2.27}$$

由独立性可以推出不相关性，但是反过来是不成立的，试看下例.

［例 7］ 设 θ 服从$[0,2\pi]$均匀分布，$\xi=\cos\theta,\eta=\cos(\theta+a)$，这里 a 是定数. 我们有

$$E\xi=\frac{1}{2\pi}\int_0^{2\pi}\cos t\,\mathrm{d}t=0,E\eta=\frac{1}{2\pi}\int_0^{2\pi}\cos(t+a)\mathrm{d}t=0$$

$$E\xi^2=\frac{1}{2\pi}\int_0^{2\pi}\cos^2 t\,\mathrm{d}t=\frac{1}{2},E\eta^2=\frac{1}{2\pi}\int_0^{2\pi}\cos^2(t+a)\mathrm{d}t=\frac{1}{2}$$

$$E\xi\eta=\frac{1}{2\pi}\int_0^{2\pi}\cos t\cos(t+a)\mathrm{d}t=\frac{1}{2}\cos a$$

因此

$$r=\cos a$$

$$\left.\begin{array}{l}\text{当 } a=0 \text{ 时}, r=1,\xi=\eta\\ \text{当 } a=\pi \text{ 时}, r=-1,\xi=-\eta\end{array}\right\}\text{存在线性关系.}$$

但是，当 $a=\frac{\pi}{2}$ 或$\frac{3\pi}{2}$时，$r=0$，这时 ξ 与 η 不相关. 不过，这时却有 $\xi^2+\eta^2=1$，因此 ξ 与 η 不独立.

这个例子给我们：(1)提供了 $r=\pm1$ 之例；(2)提供 $r=0$ 之例；(3)说明不能由不相关性推出独立性；(4)说明即使 ξ 与 η 不相关，它们之间也还是可能存在函数关系. 事实上，相关系数只是 ξ 与 η 间线性关系程度的一种量度.

不过，在一种重要的特殊场合——正态分布，独立性与不相关性却是一致的. 我们先对二元的场合来讨论这个事实.

为此，我们先求二元正态分布(3.2.22)的相关系数.

$$b_{12}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-a_1)(y-a_2)p(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \int_{-\infty}^{\infty} e^{-(y-a_2)^2/(2\sigma_2^2)} dy \int_{-\infty}^{\infty} (x-a_1)$$

$$\cdot (y-a_2) \exp\left\{-\frac{1}{2(1-r^2)}\left(\frac{x-a_1}{\sigma_1} - r\frac{y-a_2}{\sigma^2}\right)^2\right\} dx$$

作变数变换 $z=\frac{1}{\sqrt{1-r^2}}\left(\frac{x-a_1}{\sigma_1}-r\frac{y-a_2}{\sigma_2}\right)$，$t=\frac{y-a_2}{\sigma_2}$，则

$$b_{12} = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-r^2}tz + r\sigma_1\sigma_2 t^2) e^{-\frac{t^2}{2}-\frac{z^2}{2}} dz\, dt$$

$$= \frac{r\sigma_1\sigma_2}{2\pi}\int_{-\infty}^{\infty} t^2 e^{-t^2/2} dt \int_{-\infty}^{\infty} e^{-z^2/2} dz + \frac{\sigma_1\sigma_2\sqrt{1-r^2}}{2\pi}$$

$$\cdot \int_{-\infty}^{\infty} t\, e^{-t^2/2} dt \int_{-\infty}^{\infty} z\, e^{-z^2/2} dz = r\sigma_1\sigma_2$$

因此

$$r_{12} = \frac{b_{12}}{\sigma_1\sigma_2} = r \qquad (4.2.28)$$

这样说明了参数 r 是二元正态分布的相关系数.

至此，我们已经完全搞清了二元正态分布中各个参数的含义. a_1，a_2 分别是两个边际分布的数学期望，而 σ_1，σ_2，r 则以下列形式构成它的协方差矩阵

$$\begin{pmatrix} \sigma_1^2 & r\sigma_1\sigma_2 \\ r\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \qquad (4.2.29)$$

另外，在上章§2例4中已指出，二元正态分布场合，独立的充要条件是 $r=0$，这表明：

性质4 对于二元正态分布，不相关性与独立性是等价的.

在§6中，我们将把这个结果推广到多元的场合.

下面，我们给出一个边际分布是正态分布而联合分布不是多元正态分布的例子.

[例8] 令 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，$-\infty<x<\infty$

$$g(x)=\begin{cases}\cos x, & |x|<\pi \\ 0, & |x|\geqslant\pi\end{cases}$$

$$p(x,y)=\varphi(x)\varphi(y)+\frac{1}{2\pi}e^{-\pi^2}g(x)g(y),\quad -\infty<x,y<\infty$$

关于 $p(x,y)$，不难验证：(1) 是二元密度函数；(2) 边际分布都是正态分布；(3) 相关系数为 0；(4) 不独立；(5) 不是二元正态密度函数. 这些留给读者作为练习.

[在抽样调查中的应用]　抽样调查 (sample survey) 是社会经济中用得最多的统计方法. 在国外，上至总统竞选前的民意调查，下至针对家庭的家计调查，十分普遍. 最近几年，我国也已经进行了不少有效的抽样调查.

为对总体的某个指标（主要是总值、平均值、比率和百分比）进行估算，特设计某种抽样方案，以随机的方式抽取若干个体作调查，利用所得数据算出估计值，并希望给出估计的精度. 这是抽样调查的大意.

最简单也是最基本的抽样方式是所谓**简单随机抽样**，这时总体的 N 个个体中的每一个在一个大小为 n 的抽样中有同样的机会被抽到.

下面的例子可以概括水稻产量、家庭收入、收视率等等具体情况，采用的是简单随机抽样.

[例 9]　袋中有 N 张卡片，各记以数字 $Y_1, Y_2, \cdots, Y_N$，不放回地从中抽出 n 张，求其和的数学期望与方差.

[解]　取一张时，其数字的均值及方差分别为

$$\overline{Y}=\frac{1}{N}\sum_{i=1}^{N}Y_i \tag{4.2.30}$$

及

$$\sigma^2=\frac{1}{N}\sum_{i=1}^{N}(Y_i-\overline{Y})^2 \tag{4.2.31}$$

若以 η_n 记 n 张卡片的数字之和，以 $\xi_i, i=1,2\cdots,n$ 记第 i 次抽得的卡片上的数字，则

$$\eta_n=\xi_1+\xi_2+\cdots+\xi_n$$

$$P\{\xi_i=Y_l\}=\frac{1}{N},\quad l=1,2,\cdots,N,\ i=1,2,\cdots,n$$

因此

$$E\xi_i=\overline{Y},\quad D\xi_i=\sigma^2$$

所以

$$E\eta_n=E\xi_1+E\xi_2+\cdots+E\xi_n=n\overline{Y} \tag{4.2.32}$$

$$\begin{aligned} D\eta_n &= \sum_{i=1}^{n} D\xi_i+2\sum_{1\leqslant i<j\leqslant n}\operatorname{cov}(\xi_i,\xi_j) \\ &= n\sigma^2+n(n-1)\operatorname{cov}(\xi_1,\xi_2) \end{aligned} \tag{4.2.33}$$

这里用到 ξ_i 之间的对称性，也即抽签与顺序无关.

在(4.2.33)中令 $n=N$，这时 $\eta_N=Y_1+Y_2+\cdots+Y_N$ 是一个常数，因此 $D\eta_N=0$，于是

$$N\sigma^2+N(N-1)\operatorname{cov}(\xi_1,\xi_2)=0$$

所以

$$\operatorname{cov}(\xi_1,\xi_2)=-\frac{\sigma^2}{N-1} \tag{4.2.34}$$

最后得到

$$\begin{aligned} D\eta_n &= n\sigma^2-\frac{n(n-1)\sigma^2}{N-1} \\ &= \frac{n(N-n)}{N-1}\sigma^2 \end{aligned} \tag{4.2.35}$$

与有放回抽取的方差 $n\sigma^2$ 相比，多出了一个因子 $\frac{N-n}{N-1}$，称为**有限修正因子**. 当 $n=1$ 时，它等于 1；而当 $n=N$ 时，它取值为 0. 在不放回场合，添加抽取的信息量较大，即方差小，这与直观完全符合.

特别地，若取 $Y_1=Y_2=\cdots=Y_M=1$，$Y_{M+1}=\cdots=Y_N=0$，则可以得到超几何分布的均值和方差的表达式.

[现代证券组合理论]　马库威茨（Markowitz）在 50 年代引进的均值一方差模型成了现代证券组合理论的基石. 在这个理论中，假定有 n 种证券可以投资，并把它们的收益率看作是随机

变量，通常记为r_1，r_2，…，r_n，相应的均值记为$\bar{r}_1$，$\bar{r}_2$，…，$\bar{r}_n$，方差记为σ_1^2，σ_2^2，…，σ_n^2，并以ρ_{ij}记r_i与r_j的相关系数．一个相当自然的假定是：投资者都追求高收益而规避风险，也即希望有高的均值而不愿有大的方差．但是证券市场的历史记录表明，对于个别证券而言，高收益总是伴随着高风险．根本的出路在于采用证券组合，即把全部资金分散投资于各种证券．假定投资于上述n种证券的资金比例分别为w_1，w_2，…，w_n，则总的收益率为

$$r_p=\sum_{i=1}^{n}w_ir_i$$

显然，其平均收益率为

$$\bar{r}_p=Er_p=\sum_{i=1}^{n}w_i\bar{r}_i \tag{4.2.36}$$

而方差则为

$$\sigma_p^2=Dr_p=\sum_{i=1}^{n}\sum_{j=1}^{n}w_iw_j\rho_{ij}\sigma_i\sigma_j \tag{4.2.37}$$

因此寻找最优证券组合的问题化为：求投资比例w_1，w_2，…，w_n，使$\bar{r}_p$等于某个目标值而其σ_p^2达到最小，或者σ_p^2控制在一个可以接受的水平而使$\bar{r}_p$达到最大．

马库威茨模型兼顾了金融市场中收益和风险两大要素，而且形式简便，为金融学的发展开创了新局面，他也因此获得了1990年度的诺贝尔经济学奖．

四、矩

数学期望，方差，协方差是随机变量最常用的数字特征，它们都是某种矩．矩（moment）是最广泛使用的一种数字特征，在概率论和数理统计中占有重要地位．最常用的矩有两种：一种是原点矩，一种是中心矩．

定义 4.2.5 对正整数k，称

$$m_k = E\xi^k \tag{4.2.38}$$

为**k阶原点矩**．数学期望是一阶原点矩．

定义 4.2.6 对正整数k，称

$$c_k = E(\xi - E\xi)^k \tag{4.2.39}$$

为k**阶中心矩**. 方差是2阶中心矩.

由于

$$\begin{aligned} c_k = E(\xi - E\xi)^k &= \sum_{i=0}^{k} \binom{k}{i} (-E\xi)^{k-i} E\xi^i \\ &= \sum_{i=0}^{k} \binom{k}{i} (-m_1)^{k-i} m_i \end{aligned} \tag{4.2.40}$$

故中心矩可通过原点矩来表达，反之，

$$\begin{aligned} m_k &= E\xi^k = E[(\xi - m_1) + m_1]^k \\ &= \sum_{i=0}^{k} \binom{k}{i} E(\xi - m_1)^{k-i} m_1^i = \sum_{i=0}^{k} \binom{k}{i} c_{k-i} m_1^i \end{aligned} \tag{4.2.41}$$

因此当已知数学期望之后，原点矩也可以通过中心矩给出.

此外对正数p还可以定义p**阶原点绝对矩**$E|\xi|^p$及p**阶中心绝对矩**$E|\xi - E\xi|^p$，它们较少使用.

[例 10] 设ξ为正态随机变量，其密度函数为

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)}$$

因此$E\xi = 0$，故

$$m_k = c_k = E\xi^k = \int_{-\infty}^{\infty} x^k \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/(2\sigma^2)} dx \tag{4.2.42}$$

显然，k为奇数时，$c_k = 0$；k为偶数时，

$$\begin{aligned} c_k &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{x^k}{\sigma} e^{-\frac{x^2}{2\sigma^2}} dx = \sqrt{\frac{2}{\pi}} \sigma^k \cdot 2^{\frac{k-1}{2}} \int_0^{\infty} z^{\frac{k-1}{2}} e^{-z} dz \\ &= \sqrt{\frac{2}{\pi}} \sigma^k 2^{\frac{k-1}{2}} \Gamma\left(\frac{k+1}{2}\right) = \sigma^k (k-1)(k-3)\cdots 3 \cdot 1 \end{aligned} \tag{4.2.43}$$

特别地

$$c_4 = 3\sigma^4 \tag{4.2.44}$$

对于多维随机变量，可以定义各种混合矩，例如

$$E(\xi - E\xi)^k(\eta - E\eta)^l \tag{4.2.45}$$

称为 **$k+l$ 阶混合中心矩**. 协方差是二阶混合中心矩，是其中最重要的一种. 这里我们不一一赘述了.

*五、条件数学期望，最佳线性预测

在第三章 § 2 我们曾经引进了条件分布函数的概念，现在要相应地引进条件数学期望的概念，并说明它的应用.

为方便起见，我们讨论两个随机变量 ξ 与 η 的场合，假定它们具有密度函数 $p(x,y)$，并以 $p(y|x)$ 记已知 $\xi=x$ 的条件下，η 的条件密度函数，以 $p_1(x)$ 记 ξ 的密度函数.

定义 4.2.7 在 $\xi=x$ 的条件下，η 的**条件数学期望**定义为

$$E\{\eta|\xi = x\} = \int_{-\infty}^{\infty} yp(y|x)\mathrm{d}y \tag{4.2.46}$$

[例 11] 若 (ξ, η) 服从二元正态分布，则由 (3.2.29) 知

$$p(y|x) = \frac{1}{\sigma_2\sqrt{2\pi}\sqrt{1-r^2}}\exp\left\{-\frac{1}{2\sigma_2^2(1-r^2)}\cdot\left[y-\left(a_2+r\frac{\sigma_2}{\sigma_1}(x-a_1)\right)\right]^2\right\}$$

这是正态分布 $N\left(a_2+r\frac{\sigma_2}{\sigma_1}(x-a_1),\ \sigma_2^2(1-r^2)\right)$，因此

$$E\{\eta|\xi = x\} = a_2 + r\frac{\sigma_2}{\sigma_1}(x-a_1) \tag{4.2.47}$$

值得提醒的是：这时条件数学期望 $E\{\eta|\xi=x\}$ 是 x 的线性函数.

条件数学期望在**预测**问题中起重要作用. 问题这样提出：若 ξ，η 是相依的随机变量，我们要找 ξ 与 η 的函数关系，设这个关系是 $y=h(x)$，如果 $E\eta^2$ 及 $E[h(\xi)]^2$ 都存在，我们的目的是找函数 $h(x)$，使 η 与 $h(\xi)$ "尽可能靠近"，这里的 "靠近" 需要一个标准，最常用的是所谓 "最小二乘法"，这时要求使

$$E[\eta - h(\xi)]^2 \tag{4.2.48}$$

达到最小. 因为

$$E[\eta - h(\xi)]^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[y - h(x)]^2 p(x,y)\mathrm{d}x\mathrm{d}y$$
$$= \int_{-\infty}^{\infty} p_1(x)\left\{\int_{-\infty}^{\infty}[y - h(x)]^2 p(y|x)\mathrm{d}y\right\}\mathrm{d}x \tag{4.2.49}$$

由 (4.2.8) 知道，当 $h(x) = E\{\eta|\xi = x\}$ 时，$\int_{-\infty}^{\infty}[y-h(x)]^2 \cdot p(y|x)\mathrm{d}y$ 达到最小，从而使 (4.2.49) 达到最小. 即当我们观察到 $\xi = x$ 时，$E\{\eta|\xi = x\}$ 是一切对 η 的估值中均方误差最小的一个.

今后我们将称 $y = E\{\eta|\xi = x\}$ 是 η 关于 ξ 的**回归**.

通常，(ξ,η) 的联合分布函数是不知道的，或者虽然知道但是却不易算出 $E\{\eta|\xi = x\}$. 假定已知 ξ 与 η 的数学期望 μ_1，μ_2，标准差 σ_1，σ_2 及相关系数 r，这时可以降低一点要求，改为求**最佳线性预测**. 也就是说，把 $h(x)$ 限定为 x 的线性函数 $L(x) = a + bx$，求 a，b 使

$$e(a,b) = E[\eta - (a + b\xi)]^2 \tag{4.2.50}$$

达到最小.

把 $e(a,b)$ 对 a，b 求偏导数并令它们等于 0，得到

$$\begin{aligned} 2E[\eta - (a + b\xi)] &= 0 \\ 2E[(\eta - (a + b\xi))\xi] &= 0 \end{aligned} \tag{4.2.51}$$

整理后变成

$$\begin{aligned} a + b\mu_1 &= \mu_2 \\ a\mu_1 + bE\xi^2 &= E\xi\eta \end{aligned} \tag{4.2.52}$$

因此解得

$$a = \mu_2 - b\mu_1, \quad b = \frac{\operatorname{cov}(\xi,\eta)}{\sigma_1^2} = r \cdot \frac{\sigma_2}{\sigma_1} \tag{4.2.53}$$

最佳线性预测为

$$L(x) = \mu_2 + r\frac{\sigma_2}{\sigma_1}(x - \mu_1) \tag{4.2.54}$$

我们称（4.2.54）为 η 关于 ξ 的**线性回归**. 这个结果与 $E\{\eta|\xi=x\}$ 一般是不同的，但是在 (ξ,η) 是二元正态分布的场合，由（4.2.47）知两者是重合的，所以在正态分布场合，最佳预测是线性预测，这是一个十分重要的结果.

进一步，我们还可以计算最佳线性预测的均方误差.

$$\begin{aligned}E[\eta - L(\xi)]^2 &= E[\eta - \mu_2 - b(\xi - \mu_1)]^2 \\ &= \sigma_2^2 + b^2\sigma_1^2 - 2b\operatorname{cov}(\xi,\eta) \\ &= \sigma_2^2 - \frac{\operatorname{cov}^2(\xi,\eta)}{\sigma_1^2} = \sigma_2^2(1 - r^2) \qquad (4.2.55)\end{aligned}$$

因此预测误差同 η 的方差有关，也同 ξ 与 η 的相关系数有关，特别当 $|r|=1$ 时（这时 ξ 与 η 有线性关系），预测误差为 0，也就是说，可以完全准确地进行线性预测. 从这个讨论再次看出，相关系数反映了 ξ 与 η 线性联系的程度.

最佳线性预测理论中的另一个重要事实是：预测值 $\hat{\eta}=L(\xi)$ 与残差 $\eta-\hat{\eta}$ 是不相关的. 证明如下：

由（4.2.54）知

$$\hat{\eta} = L(\xi) = \mu_2 + r\frac{\sigma_2}{\sigma_1}(\xi - \mu_1) \qquad (4.2.56)$$

因此

$$E\hat{\eta} = \mu_2 \qquad (4.2.57)$$

$$E(\eta - \hat{\eta}) = 0 \qquad (4.2.58)$$

这样一来

$$\begin{aligned}\operatorname{cov}(\hat{\eta},\eta - \hat{\eta}) &= E[(\hat{\eta} - \mu_2)(\eta - \hat{\eta})] \\ &= E\left\{r\frac{\sigma_2}{\sigma_1}(\xi - \mu_1)\left[(\eta - \mu_2) - r\frac{\sigma_2}{\sigma_1}(\xi - \mu_1)\right]\right\} \\ &= r\frac{\sigma_2}{\sigma_1}\left(r\sigma_1\sigma_2 - r\frac{\sigma_2}{\sigma_1}\sigma_1^2\right) = 0 \qquad (4.2.59)\end{aligned}$$

这个事实可以解释为：残差中已不再包含对预测 η 有用的知识. 因此观察值 η 被分解为两个不相关的随机变量之和：

$$\eta = \hat{\eta} + (\eta - \hat{\eta}) \qquad (4.2.60)$$

若以 $E\{\eta|\xi\}$ 记随机变量 ξ 的如下函数：当 $\xi=x$ 时，它取值 $E\{\eta|\xi=x\}$. 这样定义的 $E\{\eta|\xi\}$ 是随机变量，对它可以求数学期望，并有下列关系式：

$$E\eta=E[E\{\eta|\xi\}] \tag{4.2.61}$$

这是条件数学期望的一个极端重要的性质，有广泛应用，在离散型及连续型随机变量的场合，读者不难对它加以证明.

*§3. 熵与信息①

一、不肯定性与熵

随机试验的主要特征是在试验之前无法肯定地知道哪一个结果将会出现，即随机试验具有一种不肯定性. 但是对于不同的随机试验，这种不肯定性的程度却可以有很大的差别. 譬如，以射击为例，若有两个射手，他们的射击情况分别以下列两个随机试验来描述：

$$甲:\begin{pmatrix}A, & \overline{A}\\ 0.5, & 0.5\end{pmatrix} \qquad 乙:\begin{pmatrix}A, & \overline{A}\\ 0.99, & 0.01\end{pmatrix}$$

这里 A 表示射中目标，$\overline{A}$ 表示未射中目标，下一行是相应的概率. 显然这两个试验的不肯定性程度很不相同，甲的不肯定性要大得多. 假如还有第三个射手，用来描述其射击水平的随机试验为

$$丙:\begin{pmatrix}A, & \overline{A}\\ 0.7, & 0.3\end{pmatrix}$$

显然应认为此试验的不肯定性程度介于上述两者之间.

因此有必要从数值上估计各种各样随机试验的不肯定性程度，即我们希望找到一个量，用它可以合理地作为不肯定性程度的度量. 这样的一个量已经被美国数学家申农（Shannon）找到.

假定我们研究的随机试验 α 只有有限个不相容的结果 A_1，

---

① 此节内容与本书其余部分基本上独立，初学时可跳过.

A_2，…，A_n，它们相应的概率为 $p(A_1)$，$p(A_2)$，…，$p(A_n)$，满足 $\sum_{i=1}^{n} p(A_i)=1$，简写如下：

$$\alpha:\begin{pmatrix} A_1, & A_2, & \cdots, & A_n \\ p(A_1), & p(A_2), & \cdots, & p(A_n) \end{pmatrix}$$

我们希望找到一个量 $H(\alpha)$ 来度量 α 的不肯定性程度. 这个量当然依赖于 $p(A_1)$，$p(A_2)$，…，$p(A_n)$，因此亦记为 $H(p(A_1), p(A_2), \cdots, p(A_n))$.

为了具体定出 $H(\alpha)$ 的表达式，我们先考察一下，究竟 $H(\alpha)$ 应满足什么要求.

首先，我们要求

(i) H 是 $p(A_i)$ 的连续函数.

这个要求相当自然，一方面 $p(A_i)$ 的微小变化当然不应引起 H 的巨大变化，同时也只有连续函数才便于数学上处理.

其次，我们考虑一种特殊的随机试验，这种试验有 n 个结果，且各个结果出现的概率均为 $\frac{1}{n}$，以后简称为**有 n 个等概结果的试验**. 在这种特殊试验中，H 当然应该只是 n 的函数，并且当 n 增大时，也即试验有更多可能的结果时，其相应的不肯定性程度也随之增加.

因此，我们对 H 提出要求：

(ii) 对有 n 个等概结果的试验，H 是 n 的单调上升函数.

对 $H(\alpha)$ 的第三个要求比较复杂，它牵涉到把一个试验分为相继的两个试验. 我们通过简单的例子来阐述其含义.

考虑有三个结果的试验

$$\alpha:\begin{pmatrix} A_1, & A_2, & A_3 \\ p_1, & p_2, & p_3 \end{pmatrix}$$

这个试验的不肯定性程度 $H(\alpha)=H(p_1,p_2,p_3)$. 为了确定到底是哪一个结果出现，我们也可以进行这样相继的两个试验：在第一个试验中，先确定到底是 A_1 出现，还是 A_2 或 A_3 出现，即进行

下列试验：

$$\alpha_1: \begin{pmatrix} A_1, & B \\ p_1, & p_2+p_3 \end{pmatrix}$$

显然 $H(\alpha_1)=H(p_1,p_2+p_3)$. 如果 A_1 出现（其概率为 p_1），则试验结果已完全确定，无须再作进一步的试验；但是如果是 B 出现（其概率为 p_2+p_3），则尚须进行如下试验，才能最后确定试验结果：

$$\alpha_2: \begin{pmatrix} A_2, & A_3 \\ \dfrac{p_2}{p_2+p_3}, & \dfrac{p_3}{p_2+p_3} \end{pmatrix}$$

这个试验的不肯定性程度为 $H(\alpha_2)=H\left(\dfrac{p_2}{p_2+p_3},\dfrac{p_3}{p_2+p_3}\right)$.

可以直接进行试验 α，以确定 A_1, A_2, A_3 中哪一个结果出现；但是若先进行试验 α_1，然后有必要时（概率 p_2+p_3）再进行 α_2，也可达到同样的目的. 因此，我们自然认为这两组试验所含的不肯定性程度是一样的，即

$$H(p_1,p_2,p_3)=H(p_1,p_2+p_3)+(p_2+p_3)H\left(\frac{p_2}{p_2+p_3},\frac{p_3}{p_2+p_3}\right)$$

这些考虑启发我们对 H 提出下列要求：

(iii) 一个试验分成相继的两个试验时，未分之前的 H 是既分之后的 H 的加权和.

条件 (i)、(ii)、(iii) 已完全确定 H 的形式. 为书写方便起见，下面简记 $p(A_i)$ 为 p_i.

定理 4.3.1（申农） 唯一满足 (i)、(ii)、(iii) 三个条件的 H 具有下列形式：

$$H=-C\sum_{i=1}^{n}p_i\log p_i \tag{4.3.1}$$

其中 C 是正常数.

为证明这个定理，要用到下述分析引理.

引理 4.3.1 若 $f(n)$ 是 n 的单调上升函数，且对一切正整数

m，n 成立

$$f(mn) = f(m) + f(n) \tag{4.3.2}$$

则

$$f(n) = C\log n$$

其中 C 是一个正常数.

［证明］ 由（4.3.2）可得 $f(1) = 0$，所以对其他正整数 m，有 $f(m) > 0$．另一方面，

$$f(n^2) = f(n) + f(n) = 2f(n)$$
$$f(n^3) = f(n^2) + f(n) = 3f(n)$$

一般地

$$f(n^k) = kf(n) \tag{4.3.3}$$

若 n，m 是两个任意的正整数，$m \neq 1$，选任意大的正整数 k，再取正整数 l，使

$$m^l \leqslant n^k < m^{l+1} \tag{4.3.4}$$

由函数的单调性

$$f(m^l) \leqslant f(n^k) < f(m^{l+1})$$

由（4.3.3）得

$$lf(m) \leqslant kf(n) < (l+1)f(m)$$

因此

$$\frac{l}{k} \leqslant \frac{f(n)}{f(m)} < \frac{l+1}{k}$$

对（4.3.4）取对数得

$$l\log m \leqslant k\log n < (l+1)\log m$$

因此也有

$$\frac{l}{k} \leqslant \frac{\log n}{\log m} < \frac{l+1}{k}$$

这样一来

$$\left| \frac{f(n)}{f(m)} - \frac{\log n}{\log m} \right| < \frac{1}{k}$$

上式对任意大的 k 都成立，因此

$$\frac{f(n)}{f(m)}=\frac{\log n}{\log m}$$

由于 m，n 的任意性即知

$$f(n)=C\log n$$

其中 C 是常数，再由 $f(n)$ 是 n 的上升函数，可知 C 是正的.

现在可以来证明定理了.

[证明]　首先，记 $H\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right)=f(n)$，按条件(ii)知 $f(n)$ 是 n 的单调上升函数. 对有 mn 个等概结果的试验，可以把它分解为 m 个有 n 个等概结果的试验，因此由条件（iii）知

$$f(mn)=f(m)+m\cdot\frac{1}{m}f(n)=f(m)+f(n)$$

利用引理 4.3.1 立刻得到

$$H\left(\frac{1}{n},\frac{1}{n},\cdots,\frac{1}{n}\right)=C\log n$$

其次，当 p_1，p_2，…，p_n 是有理数时，不妨记 $p_i=\frac{n_i}{\sum_{i=1}^{n}n_i}$，考虑一个有 $\sum_{i=1}^{n}n_i$ 个等概结果的试验，而这个试验又可以看作两个相继的试验，其中第一个试验以概率 p_i 出现结果 A_i；而第二个试验，则是在出现结果 A_i 的基础上，考察它是出现 n_i 个等概结果中的哪一个，因此按条件（iii）应有

$$C\log\sum_{i=1}^{n}n_i=H(p_1,p_2,\cdots,p_n)+C\sum_{i=1}^{n}p_i\log n_i$$

于是

$$\begin{aligned}H(p_1,p_2,\cdots,p_n)&=C\left[\log\sum_{i=1}^{n}n_i-\sum_{i=1}^{n}p_i\log n_i\right]\\&=C\left[\sum_{i=1}^{n}p_i\left(\log\sum_{j=1}^{n}n_j-\log n_i\right)\right]\\&=-C\sum_{i=1}^{n}p_i\log p_i\end{aligned}$$

.

最后，对任意的实数 p_1，p_2，…，p_n，可用有理数来逼近它，但按条件 (i) 知 H 是各自变量的连续函数，因此上述表达式仍然成立，从而完成了定理证明.

定理中的系数 C，可以根据方便选择，它取决于度量单位. 常用的度量单位有二进制单位及十进制单位，前者对数的底取为 2，后者用常用对数.

在 (4.3.1) 中，若 $p_i=0$，则相应的项 $p_i\log p_i$ 定义为零，因此在试验中增减零概率结果不影响不肯定性，这是很自然的.
以后将称

$$H(\alpha)=-\sum_{i=1}^{n}p(A_i)\log p(A_i) \tag{4.3.5}$$

为试验 α 的**熵** (**entropy**).

从下面例子可以看出，熵确实可以度量试验的不肯定性程度.

[例 1] 计算本节开始时射击例子中射手甲，乙，丙射击试验相应的熵.

$$H_{甲}=-\frac{1}{2}\lg\frac{1}{2}-\frac{1}{2}\lg\frac{1}{2}=\lg 2=0.3010$$

$$H_{乙}=-0.99\lg 0.99-0.01\lg 0.01=0.0243$$

$$H_{丙}=-0.7\lg 0.7-0.3\lg 0.3=0.2653$$

这里均用十进制单位. 甲的熵最大，乙最小，丙介于其中，与直观完全符合.

[例 2] 英文字母出现的熵.

熵的概念的引入与通信理论的发展密切相关，后者需要解决最有效而无错误地传递消息这一任务. 用文字给出的消息通常经过编码变成某种信号，经信道传输，到达接收地点，再经译码还原. 当然编码的优劣将直接影响通信的效率. 最简单的办法是把每一个字母都转换成相同长度的码子，但正如本书一开头就看到的那样，在英文中不同字母出现的频率很不相同. 因此更有效的办法当然是把最常出现的字母编成较短的码子，而把不常出现的

字母编成较长的码子. 这个问题与英文字母出现的熵有直接关系.

假如把 26 个字母连同分隔用的空格共 27 个符号，看作是等可能出现的，则相应的试验（接收到一个符号，要判断它是哪个字母或空格）的熵为

$$H_0 = \lg 27 = 1.4314$$

但是接收到一个符号，它是空格的可能性比是 z 的可能性要大 200 倍，也即不肯定性程度似乎不应该有这么大. 事实上，如果我们考虑到不同字母及空格的出现概率（数值见第一章 § 1），则

$$H_1 = -\sum_{i=1}^{27} p_i \lg p_i = 1.213$$

也就是说不肯定性要小不少. 当然如果再注意到英文中前后字母间的联系，例如 q 总是接着 u，则这种不肯定性还要小得多.

二、熵的基本性质

下面讨论用（4.3.5）定义的熵的若干基本性质，通过这些研究将进一步看到熵作为不肯定性程度度量的合理性.

首先考察一下函数 $\varphi(x) = -x\log x$ 的性质，显然对于 $x>0$，均有 $\varphi''(x) < 0$，因此 $\varphi(x)$ 是 $(0, \infty)$ 上的上凸函数，即对于任意的 $p>0$，$q>0$，且 $p+q=1$，不等式

$$p\varphi(x_1) + q\varphi(x_2) < \varphi(px_1 + qx_2) \tag{4.3.6}$$

对一切 $0<x_1<x_2<\infty$ 成立.

一般地，不难用归纳法证明如下分析引理.

引理 4.3.2（Jensen 不等式） 设 $\varphi(x)$ 是 $[a, b]$ 上的上凸函数，而 $x_1, x_2, \cdots, x_n$ 是 $[a, b]$ 中的任意点，$\lambda_1, \lambda_2, \cdots, \lambda_n$ 是和为 1 的正数，则

$$\sum_{i=1}^{n} \lambda_i \varphi(x_i) \leqslant \varphi\Big(\sum_{i=1}^{n} \lambda_i x_i\Big) \tag{4.3.7}$$

等号成立当且仅当诸 x_i 相等.

下面证明熵的若干性质.

1. 当且仅当 $p(A_i)$，$i=1, 2, \cdots, n$ 之中的一个等于1时，熵 $H=0$，其他情况下，熵恒为正.

2. 在有 n 个可能结果的试验中，等概试验具有最大熵，其值为 $\log n$.

[证明] 在引理4.3.2中取 $\varphi(x)=-x\log x$，$x_i=p(A_i)$，$\lambda_i=\frac{1}{n}$，代入（4.3.7）式，得到

$$\frac{-1}{n}\sum_{i=1}^{n}p(A_i)\log p(A_i)\leqslant\frac{-1}{n}\log\frac{1}{n}$$

即

$$H(p(A_1),\cdots,p(A_n))\leqslant\log n=H\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$$

下面考虑两个试验 α 及 β，设它们的结果及概率如下

$$\alpha:\begin{Bmatrix}A_1, & \cdots, & A_m\\ p(A_1), & \cdots, & p(A_m)\end{Bmatrix}\qquad\beta:\begin{Bmatrix}B_1, & \cdots, & B_n\\ p(B_1), & \cdots, & p(B_n)\end{Bmatrix}$$

又以 $\alpha\beta$ 记这两个试验联合起来所构成的新试验，于是试验 $\alpha\beta$ 的可能结果为 A_kB_l，$k=1, 2, \cdots, m$，$l=1, 2, \cdots, n$，相应的概率为 $p(A_kB_l)$. 按定义

$$H(\alpha\beta)=-\sum_{k,l}p(A_kB_l)\log p(A_kB_l)\qquad(4.3.8)$$

3. 若试验 α 与试验 β 独立，则

$$H(\alpha\beta)=H(\alpha)+H(\beta)\qquad(4.3.9)$$

[证明] 在这种场合 $p(A_kB_l)=p(A_k)p(B_l)$，因此

$$\begin{aligned}H(\alpha\beta)&=-\sum_{k,l}p(A_k)p(B_l)\log p(A_k)p(B_l)\\&=-\sum_{k,l}p(A_k)p(B_l)[\log p(A_k)+\log p(B_l)]\\&=H(\alpha)+H(\beta)\end{aligned}$$

三、条件熵与信息量

为了进一步研究熵的性质，需要引进条件熵的概念. 设 α，β

是前述两个试验，以 $p(B_l|A_k)$ 记试验 α 出现结果 A_k 的条件下，试验 β 出现结果 B_l 的概率，则

$$H_{A_k}(\beta)=-\sum_{l=1}^{n}p(B_l|A_k)\log p(B_l|A_k) \qquad (4.3.10)$$

是在试验 α 出现 A_k 的条件下，试验 β 的熵.

我们称平均值

$$H_\alpha(\beta)=\sum_{k=1}^{m}p(A_k)H_{A_k}(\beta) \qquad (4.3.11)$$

为在试验 α 实现的条件下试验 β 的**条件熵**.

下面指出 $H_\alpha(\beta)$ 的某些重要性质.

性质 1 $H(\alpha\beta)=H(\alpha)+H_\alpha(\beta).$ (4.3.12)

[证明]

$$\begin{aligned}H(\alpha\beta)&=-\sum_{k,l}p(A_kB_l)\log p(A_kB_l)\\&=-\sum_{k,l}p(A_k)p(B_l|A_k)[\log p(A_k)+\log p(B_l|A_k)]\\&=-\sum_{k}p(A_k)\log p(A_k)\sum_{l}p(B_l|A_k)\\&\quad-\sum_{k}p(A_k)\sum_{l}p(B_l|A_k)\log p(B_l|A_k)\\&=H(\alpha)+H_\alpha(\beta)\end{aligned}$$

特别当 α 与 β 独立时，$H_\alpha(\beta)=H(\beta)$，此时 (4.3.12) 化为 (4.3.9) 式.

同理，

$$H(\alpha\beta)=H(\beta)+H_\beta(\alpha)$$

这个性质称为**熵的加法法则**. 推导熵的表达式时的条件 (iii)，事实上是加法法则的另一种表述.

性质 2 $H_\alpha(\beta)$ 是非负的. 又若所有的 $p(A_i)>0$，则当且仅当 $H_{A_i}(\beta)=0(i=1,\cdots,m)$ 时，$H_\alpha(\beta)=0$ 才成立，此时还有 $H(\alpha\beta)=H(\alpha)$.

这些性质用条件熵的定义立即得到. 后一结论说明，只有当试验 α 的任何结果都使试验 β 的不肯定性完全消除时，才有

$H_\alpha(\beta)=0$，此时 α 的结果完全决定了 β 的结果.

3. $H_\alpha(\beta)\leqslant H(\beta)$. (4.3.13)

［证明］ 在引理 4.3.2 中，取 $\varphi(x)=-x\log x$，$\lambda_i=p(A_i)$，$x_i=p(B_k|A_i)$，则

$$
\begin{aligned}
&-\sum_{i=1}^{m}p(A_i)p(B_k|A_i)\log p(B_k|A_i)\\
&\leqslant-\Big[\sum_{i=1}^{m}p(A_i)p(B_k|A_i)\Big]\log\Big[\sum_{i=1}^{m}p(A_i)p(B_k|A_i)\Big]\\
&=-p(B_k)\log p(B_k)
\end{aligned}
$$

两边对 k 求和即得

$$
-\sum_{i=1}^{m}p(A_i)\sum_{k=1}^{n}p(B_k|A_i)\log p(B_k|A_i)\leqslant-\sum_{k=1}^{n}p(B_k)\log p(B_k)
$$

此即

$$
H_\alpha(\beta)\leqslant H(\beta)
$$

此外，显然有

$$
H(\alpha\beta)=H(\alpha)+H_\alpha(\beta)\leqslant H(\alpha)+H(\beta)
$$

关系式(4.3.13)的涵义不难理解，因为进行试验 α 之后，一般对试验 β 的结果会增加了解，从而消除了部分不肯定性，只有当 α 与 β 独立时，$H_\alpha(\beta)=H(\beta)$，此时 α 的结果无助于减少 β 的不肯定性. 因此量 $H(\beta)-H_\alpha(\beta)$ 是作了辅助试验 α 之后试验 β 不肯定性的减少量，即是由于试验 α 的进行而得到的有关试验 β 的信息.

记

$$
I(\alpha,\beta)=H(\beta)-H_\alpha(\beta) \tag{4.3.14}
$$

并称之为含在试验 α 中的有关试验 β 的**信息量**.

因为

$$
H(\alpha\beta)=H(\alpha)+H_\alpha(\beta)=H(\beta)+H_\beta(\alpha)
$$

所以

$$
H(\alpha)-H_\beta(\alpha)=H(\beta)-H_\alpha(\beta)
$$

即

$$I(\beta,\alpha) = I(\alpha,\beta)$$

因此 β 中含有 α 的信息量与 α 中含有 β 的信息量相等.

显然

$$0 \leqslant I(\alpha,\beta) \leqslant H(\alpha), \quad 0 \leqslant I(\alpha,\beta) \leqslant H(\beta) \quad (4.3.15)$$

当且仅当 α 与 β 独立时，才有 $I(\alpha,\beta) = 0$ ，此时一个试验不含有另一个试验的任何信息量. 另一个极端情形是当 α 的结果完全决定 β 的结果，此时 $H_\alpha(\beta) = 0$ ，从而 $I(\alpha,\beta) = H(\beta)$. 特别地，$I(\beta,\beta) = H(\beta)$ ，这就是说，包含在试验 β 中有关试验 β 的信息量等于 β 的熵，因而熵也是信息量.

四、连续型分布的熵

对于具有密度函数的连续型分布，可以类似地定义它的熵. 设随机变量 α 及 β 的密度函数分别为 $p(x)$ 及 $q(y)$ ，它们的联合密度函数为 $f(x,y)$. 一种比较显然的定义熵的办法是仿照离散型场合，定义

$$H(\alpha) = -\int_{-\infty}^{\infty} p(x)\log p(x)\mathrm{d}x \quad (4.3.16)$$

$$H(\alpha\beta) = -\iint f(x,y)\log f(x,y)\mathrm{d}x\mathrm{d}y \quad (4.3.17)$$

和

$$H_\alpha(\beta) = -\iint f(x,y)\log \frac{f(x,y)}{p(x)}\mathrm{d}x\mathrm{d}y \quad (4.3.18)$$

$$H_\beta(\alpha) = -\iint f(x,y)\log \frac{f(x,y)}{q(y)}\mathrm{d}x\mathrm{d}y \quad (4.3.19)$$

这样定义的熵及条件熵具有许多离散型分布的熵的性质. 简单罗列如下：

性质 1 若 α 限制在 V 中变化，则 V 中的均匀分布有最大熵，其值等于 $\log|V|$，此处 $|V|$ 是 V 的测度.

性质 2 $H(\alpha\beta) = H(\alpha) + H_\alpha(\beta) = H(\beta) + H_\beta(\alpha)$，

而且

$$H_\alpha(\beta)\leqslant H(\beta)$$

因而

$$H(\alpha\beta)\leqslant H(\alpha)+H(\beta)$$

这两个等号成立的充要条件是 α 与 β 独立.

第一个性质说明，从熵的观点看，均匀分布具有特殊地位. 令人惊奇的是，常见的另外两个分布——正态分布与指数分布，从熵的观点看，也具有特殊地位.

性质 3 设 $p(x)$ 是一元密度函数，其标准差为 σ，则当 $p(x)$ 为正态分布时其熵最大，其值等于 $\log\sqrt{2\pi e}\sigma$（对数以 e 为底）.

［证明］ 不失一般性，设其数学期望为零，这时要求 $p(x)$ 满足约束条件

$$\int p(x)\mathrm{d}x=1$$

及

$$\sigma^2=\int x^2p(x)\mathrm{d}x$$

又使

$$H(x)=-\int p(x)\log p(x)\mathrm{d}x$$

达到最大. 根据变分法，这相当于要求

$$\int[-p(x)\log p(x)+\lambda p(x)+\mu x^2p(x)]\mathrm{d}x$$

达到极大，即

$$-1-\log p(x)+\lambda+\mu x^2=0$$

选取常数使其满足约束条件，即得

$$p(x)=\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}$$

此时

$$H(x)=\int\frac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-x^2/(2\sigma^2)}[\log\sqrt{2\pi}\sigma+\frac{x^2}{2\sigma^2}]\mathrm{d}x$$

$$=\log\sqrt{2\pi}\sigma+\frac{1}{2}=\log\sqrt{2\pi e}\sigma$$

在 n 元场合，假定分布的协方差矩阵固定为 $\boldsymbol{B}=(b_{ij})$，则

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}(\det\boldsymbol{B})^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}\boldsymbol{x}\boldsymbol{B}^{-1}\boldsymbol{x}'\right\},\boldsymbol{x}\in\mathbf{R}^n$$

达到最大熵 $\log(2\pi e)^{n/2}(\det\boldsymbol{B})^{\frac{1}{2}}$.

性质 4 若密度函数 $p(x)$ 当 $x\leqslant 0$ 时等于 0，并且其均值为 a，则指数分布

$$p(x)=\frac{1}{a}e^{-x/a}$$

达到最大熵，其值为 $\log ea$.

证明方法与推导正态分布时完全一样，不再写出.

连续熵的性质也有与离散熵不同的，特别是它的数值会因坐标系的改变而改变. 因此还存在着别种关于连续熵的定义，不过这里不准备再深入讨论了.

熵是一门新兴学科——**信息论**中的基本概念. 它的引入使得人们能对随机现象的不肯定性进行度量，是具有重大意义的.

*§4. 母 函 数

一、整值随机变量与母函数的定义

在离散型随机变量中，那些只取非负整数值 0,1,2,…的占有重要的地位. 事实上，我们所遇到的离散型分布如二项分布，超几何分布，泊松分布，几何分布，巴斯卡分布，负二项分布等都是取非负整数值的.

我们称取非负整数值的随机变量为**整值随机变量**. 对于整值随机变量，有一种处理方法很便于应用，这就是母函数法.

若随机变量 ξ 取非负整数值，且相应的分布列为

$$\begin{pmatrix} 0 & 1 & 2\cdots \\ p_0 & p_1 & p_2\cdots \end{pmatrix} \tag{4.4.1}$$

则

$$P(s)=\sum_{k=0}^{\infty}p_k s^k \tag{4.4.2}$$

称为 ξ 的**母函数**(generating function)，即

$$P(s)=Es^{\xi} \tag{4.4.3}$$

因为母函数由分布列完全决定，因此亦称它为该概率分布的母函数．由于

$$\sum_{k=0}^{\infty}p_k=1 \tag{4.4.4}$$

由幂级数的收敛性知道 $P(s)$ 至少在 $|s|\leqslant 1$ 一致收敛且绝对收敛．因此母函数对任何整值随机变量都存在．

对于任一数列 $\{a_n\}$ 也可定义 $\sum_{n=0}^{\infty}a_n s^n$ 为其母函数，但我们以后只讨论概率分布对应的母函数．

母函数在19世纪初被拉普拉斯引进，它是在概率论中第一个被系统地应用的变换法，对后来在概率论中引进其他更有用的变换——如下节要介绍的特征函数——有启发作用．本书把变换法的重点放在特征函数法上，至于为什么要单独介绍母函数，首先是由于它比较简单，在整值随机变量场合很有用，可以作为特征函数的前导；其次，在随机过程中要用到有关结果；最后还由于从母函数法发展起来的 Z 变换法已成为解决许多问题的重要工具．但是，跳过本节母函数的内容，并不太影响本书以后章节的学习．

下面求几种分布的母函数．

[例 1]　二项分布

$$P(s)=\sum_{k=0}^{n}\binom{n}{k}p^k q^{n-k}s^k=(q+ps)^n \tag{4.4.5}$$

[例 2]　超几何分布

$$P(s)=\sum_{k=0}^{n}\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}s^k \tag{4.4.6}$$

这是**超几何级数**，是一种特殊函数，处理起来不大方便，在概率论中也很少用，由于超几何分布的名称来自于它，因此我们顺便提及.

[例 3]　泊松分布

$$P(s)=\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}e^{-\lambda}s^k=e^{-\lambda}\cdot e^{\lambda s}=e^{\lambda(s-1)} \tag{4.4.7}$$

[例 4]　几何分布

$$P(s)=\sum_{k=1}^{\infty}q^{k-1}ps^k=ps\sum_{k=1}^{\infty}(qs)^{k-1}=\frac{ps}{1-qs} \tag{4.4.8}$$

二、母函数的性质

1. 唯一性. 由分布列 (4.4.1) 用 (4.4.2) 定义母函数，这显然是唯一确定的；下面证明，由母函数也能唯一确定分布列.

设概率分布$\{p_k\}$及$\{q_k\}$分别具有母函数$P(s)$及$Q(s)$，且$P(s)=Q(s)$，因为$P(s)$及$Q(s)$都是幂级数，且当$|s|\leqslant 1$时收敛，对$P(s)$及$Q(s)$求导k次，并令$s=0$，则得

$$k!\ p_k=P^{(k)}(0)=Q^{(k)}(0)=k!\ q_k$$

因此$p_k=q_k$，$k=0$，1，2，…即两个概率分布一样.

这样一来，概率分布与母函数是一一对应的，因而对于概率分布的许多研究可以化为对所对应的母函数的研究，因为母函数是幂级数，具有许多良好的性质，便于处理，所以母函数是研究整值随机变量的有效工具.

2. 母函数与数字特征. 母函数的应用之一是利用它能求得概率分布的数字特征，若

$$P(s)=\sum_{k=0}^{\infty}p_ks^k$$

即

$$P'(s)=\sum_{k=1}^{\infty}kp_ks^{k-1},P''(s)=\sum_{k=2}^{\infty}k(k-1)p_ks^{k-2}$$

这两个级数至少在 $|s|<1$ 是收敛的.

当数学期望 $\sum_{k=1}^{\infty}kp_k$ 存在时,

$$P'(1)=\sum_{k=1}^{\infty}kp_k=E\xi \tag{4.4.9}$$

当数学期望 $\sum_{k=1}^{\infty}kp_k=\infty$ 时, $\lim_{s\to1}P'(s)=\infty$.

同样,当方差 $D\xi$ 存在时,

$$E[\xi(\xi-1)]=\sum_{k=2}^{\infty}k(k-1)p_k=p''(1)$$

故

$$D\xi=E\xi^2-(E\xi)^2=P''(1)+P'(1)-[P'(1)]^2 \tag{4.4.10}$$

公式(4.4.9)及(4.4.10)是计算数学期望及方差的简便公式.

[例 5] 二项分布:母函数为 $P(s)=(q+ps)^n$.

$$E\xi=P'(1)=n(q+ps)^{n-1}p|_{s=1}=np$$

$$P''(1)=n(n-1)(q+ps)^{n-2}p^2|_{s=1}=n(n-1)p^2$$

$$D\xi=n^2p^2-np^2+np-n^2p^2=npq$$

这些结果在§1与§2中曾直接计算过.

[例 6] 泊松分布:母函数为 $P(s)=e^{\lambda(s-1)}$.

$$E\xi=P'(1)=e^{\lambda(s-1)}\cdot\lambda|_{s=1}=\lambda$$

$$P''(1)=e^{\lambda(s-1)}\cdot\lambda^2|_{s=1}=\lambda^2$$

$$D\xi=\lambda^2+\lambda-\lambda^2=\lambda$$

这些结果在§1与§2中直接计算过，这里的计算较方便.

三、独立随机变量和的母函数

若随机变量 ξ 与 η 相互独立,它们都是整值随机变量,概率分布分别为 $\{a_k\}$ 及 $\{b_k\}$，而相应的母函数为 $A(s)$ 及 $B(s)$，下面计算

随机变量 $\zeta=\xi+\eta$ 的概率分布. 显然 ζ 也是整值随机变量，若记 $c_r=P\{\zeta=r\}$,则

$$c_r=a_0b_r+a_1b_{r-1}+\cdots+a_rb_0$$

这就是离散褶积公式（3.3.5).

记

$$C(s)=\sum_{r=0}^{\infty}c_rs^r$$

利用母函数在 $|s|\leqslant 1$ 的一致收敛性及绝对收敛性，

$$\begin{aligned}A(s)B(s)&=\sum_{k=0}^{\infty}a_ks^k\cdot\sum_{l=0}^{\infty}b_ls^l=\sum_{k,l}a_kb_ls^{k+l}\\&=\sum_{r=0}^{\infty}\Big(\sum_{k=0}^{r}a_kb_{r-k}\Big)s^r=\sum_{r=0}^{\infty}c_rs^r\end{aligned}$$

因此

$$C(s)=A(s)B(s)$$

即两个独立随机变量之和的母函数是这两个随机变量的母函数的乘积. 这是一个相当重要的性质，由于母函数具有这个性质，因此在研究独立随机变量和的问题时，母函数很适用.

容易把上面结果推广到 n 个独立整值随机变量之和的场合，若随机变量 ξ_1，ξ_2，…，ξ_n 相互独立，它们的母函数分别为$P_1(s)$，$P_2(s)$,…,$P_n(s)$，则 $\eta=\xi_1+\xi_2+\cdots+\xi_n$ 的母函数为

$$P(s)=P_1(s)P_2(s)\cdots P_n(s) \qquad (4.4.11)$$

特别当 ξ_i 有相同概率分布的场合,$P_i(s)=P_1(s)$,这时

$$P(s)=[P_1(s)]^n \qquad (4.4.12)$$

［例 7］ 二项分布的母函数：在成功概率为 p 的 n 次伯努利试验中，若令

$$\xi_i=\begin{cases}1, & \text{在第 } i \text{ 次试验中 } A \text{ 出现}\\0, & \text{在第 } i \text{ 次试验中 } A \text{ 不出现}\end{cases}$$

则 ξ_1，ξ_2，…，ξ_n 相互独立，而且 $\eta=\xi_1+\xi_2+\cdots+\xi_n$ 服从二项分布. ξ_i 的母函数为 $(q+ps)$，由 (4.4.12)，η 的母函数为

$$P(s)=(q+ps)^n$$

这与例 1 直接计算的结果相同.

[例 8]　掷 5 颗骰子，求所得总和为 15 的概率

[解]　若以 ξ_i 记第 i 颗掷出的数字，则总和 $\eta=\xi_1+\xi_2+\cdots+\xi_5$，$\xi_i$ 的母函数为

$$P_i(s)=\frac{1}{6}(s+s^2+s^3+s^4+s^5+s^6)$$

显然 ξ_1，ξ_2，…，ξ_5 是相互独立的，因此 η 的母函数为

$$P(s)=\frac{1}{6^5}(s+s^2+\cdots+s^6)^5$$

所求的概率 $P\{\eta=15\}$ 是 $P(s)$ 展开式中 s^{15} 项的系数. 由于

$$\begin{aligned}P(s)&=\frac{s^5}{6^5}(1+s+\cdots+s^5)^5=\frac{s^5}{6^5}\left(\frac{1-s^6}{1-s}\right)^5\\&=\frac{s^5}{6^5}(1-s^6)^5(1-s)^{-5}\\&=\frac{s^5}{6^5}(1-5s^6+10s^{12}+\cdots-s^{30})\\&\quad\cdot\left[\sum_{k=0}^{\infty}\binom{-5}{k}(-s)^k\right]\\&=\frac{s^5}{6^5}(1-5s^6+\cdots)\\&\quad\cdot\left[\sum_{k=0}^{\infty}(-1)^k\binom{5+k-1}{k}(-1)^k s^k\right]\\&=\frac{s^5}{6^5}(1-5s^6+\cdots)\left(\sum_{k=0}^{\infty}\binom{k+4}{k}s^k\right)\end{aligned}$$

故

$$P\{\eta=15\}=\frac{1}{6^5}\left[1\times\binom{14}{10}-5\times\binom{8}{4}\right]=\frac{651}{6^5}$$

掷骰子问题在概率论发展的早期一直占有显著地位，这里用母函数法给予统一处理.

四、随机个随机变量之和的母函数

若 ξ_1，ξ_2，…，ξ_n，…是一串相互独立具有相同概率分布的整

值随机变量,$P\{\xi_i=j\}=f_j$，其母函数为

$$F(s)=\sum_{j=0}^{\infty}f_js^j$$

随机变量 ν 是取正整数值的，且 $P\{\nu=n\}=g_n$，其母函数为

$$G(s)=\sum_{n=1}^{\infty}g_ns^n$$

若 $\{\xi_n\}$ 与 ν 独立，考虑和 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$，记

$$P\{\eta=i\}=h_i$$

我们来求 η 的母函数

$$H(s)=\sum_{i=0}^{\infty}h_is^i$$

利用全概率公式及$\{\xi_n\}$与 ν 的独立性

$$\begin{aligned}h_i&=P\{\eta=i\}=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\eta=i\,|\,\nu=n\}\\&=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\xi_1+\xi_2+\cdots+\xi_n=i\,|\,\nu=n\}\\&=\sum_{n=1}^{\infty}P\{\nu=n\}P\{\xi_1+\xi_2+\cdots+\xi_n=i\}\end{aligned}$$

由于 $\xi_1+\xi_2+\cdots+\xi_n$ 为 n 个相互独立相同分布的随机变量之和,故其母函数

$$\sum_{i=0}^{\infty}P\{\xi_1+\cdots+\xi_n=i\}s^i=[F(s)]^n$$

因此

$$\begin{aligned}H(s)&=\sum_{i=0}^{\infty}h_is^i=\sum_{n=1}^{\infty}P\{\nu=n\}\sum_{i=0}^{\infty}P\{\xi_1+\cdots+\xi_n=i\}s^i\\&=\sum_{n=1}^{\infty}g_n[F(s)]^n=G[F(s)]\end{aligned}\qquad(4.4.13)$$

从（4.4.13）可以看到，随机个相互独立相同分布的随机变量之和的母函数是原来两个母函数的复合.

由于

$$H'(s)=G'[F(s)]\cdot F'(s)$$

因此当 $E\xi_i$ 及 $E\nu$ 存在时，在上式中令 $s=1$，得到

$$E\eta=E\nu\cdot E\xi_i \tag{4.4.14}$$

这个公式很直观，也有不少用处.

［复合泊松分布］ 在上述讨论中，若 ν 服从参数为 λ 的泊松分布，则

$$G(s)=\mathrm{e}^{\lambda(s-1)}$$

因此 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$ 的母函数为

$$H(s)=\mathrm{e}^{\lambda(F(s)-1)} \tag{4.4.15}$$

以(4.4.15)为母函数的概率分布称为**复合泊松分布**.这种分布也很有用.

特别当 $F(s)=q+ps$ 时，

$$H(s)=\mathrm{e}^{\lambda p(s-1)}$$

［例 9］ 若每条蚕的产卵数服从泊松分布，参数为 λ，而每个卵变为成虫的概率为 p，且各卵是否变为成虫彼此间没有关系，求每条蚕养活 k 只小蚕的概率。

［解］ 这是第二章习题.现在利用上面结果立刻知道小蚕数服从参数为 λp 的泊松分布，因此所求概率等于

$$\frac{(\lambda p)^k}{k!}\mathrm{e}^{-\lambda p}$$

［例 10］ 观察资料表明，天空中星体数服从泊松分布，其参数为 λV，这里 V 是被观察区域的体积. 若每个星球上有生命存在的概率为 p，则在体积为 V 的宇宙空间中有生命存在的星球数服从参数为 λpV 的泊松分布.

［保险中的索赔模型］ 在非寿命保险中通常采用如下索赔模型：

$$\eta=\xi_1+\xi_2+\cdots+\xi_\mu \tag{4.4.16}$$

这里 ξ_i，$i=1$，2，…为独立同分布随机变量，表示第 i 次索赔数额，而总索赔次数则是随机变量 μ，因此表现为随机个独立随机变量之和的形式。更一般的索赔模型则还假定 μ 也是随机个随机变

量之和：

$$\mu=\theta_1+\theta_2+\cdots+\theta_v \tag{4.4.17}$$

例如一次事故引起随机个索赔，而事故次数也是随机的. 在这类模型中通常假定事故发生服从泊松分布.

§5. 特征函数

一、定义

数字特征只反映了概率分布的某些侧面，一般并不能通过它们来完全确定分布函数，本节将要引进的特征函数，既能完全决定分布函数而又具有良好的分析性质.

为了定义特征函数，我们需要稍为拓广一下随机变量的概念，引进复随机变量.

定义 4.5.1 如果 ξ 与 η 都是概率空间 $(\Omega, \mathscr{F}, P)$ 上的实值随机变量，则称 $\zeta=\xi+\mathrm{i}\eta$ 为**复随机变量**.

从定义知道，对复随机变量的研究本质上是对二维随机向量的研究. 这里举一个例子：如果二维向量 (ξ_1, η_1) 与 (ξ_2, η_2) 是独立的，则我们称复随机变量 $\zeta_1=\xi_1+\mathrm{i}\eta_1$ 与 $\zeta_2=\xi_2+\mathrm{i}\eta_2$ 是独立的.

定义一个复随机变量 $\zeta=\xi+\mathrm{i}\eta$ 的数学期望为

$$E\zeta=E\xi+\mathrm{i}E\eta$$

对复随机变量也可以平行于实随机变量建立起一系列结果. 例如，若 ζ_1，ζ_2，…，ζ_n 是相互独立的，则

$$E\zeta_1\zeta_2\cdots\zeta_n=E\zeta_1E\zeta_2\cdots E\zeta_n$$

又如，若 $g(x)$ 是一个一元博雷尔可测函数，而 $\eta=g(\xi)$，则

$$Ee^{\mathrm{i}t\eta}=Ee^{\mathrm{i}tg(\xi)}=\int_{-\infty}^{\infty}e^{\mathrm{i}tg(x)}\mathrm{d}F_\xi(x) \tag{4.5.1}$$

这里常用欧拉公式 $e^{\mathrm{i}t\eta}=\cos t\eta+\mathrm{i}\sin t\eta$.

以后将随时引用这类结果而不再加以说明.

下面引进随机变量 ξ 的特征函数.

定义 4.5.2 若随机变量 ξ 的分布函数为 $F_\xi(x)$，则称

$$f_\xi(t) = Ee^{it\xi} = \int_{-\infty}^{\infty} e^{itx} dF_\xi(x) \tag{4.5.2}$$

为 ξ 的**特征函数** (characteristic function).

特征函数是一个实变量的复值函数，由于 $|e^{itx}|=1$，所以它对一切实数 t 都有意义.

显然特征函数只与分布函数有关，因此亦称某一分布函数的特征函数.

对于离散型随机变量，若其分布列为

$$\begin{pmatrix} x_1 & x_2 \cdots & x_n \cdots \\ p_1 & p_2 \cdots & p_n \cdots \end{pmatrix}$$

则其特征函数为

$$f(t) = \sum_{j=1}^{\infty} p_j e^{itx_j} \tag{4.5.3}$$

特别地，对于整值随机变量，若其母函数为 $P(s)$，则 $f(t)=P(e^{it})$.

对于连续型随机变量，若其分布密度函数为 $p(x)$，则其特征函数为

$$f(t) = \int_{-\infty}^{\infty} e^{itx} p(x) dx \tag{4.5.4}$$

这时，特征函数是密度函数 $p(x)$ 的傅里叶(Fourier)变换.

一般情况下的特征函数可以看作是这种傅里叶变换的推广. 傅里叶分析是数学中一种非常有力的工具，它在许多数学分支中都起了重大作用，以后我们将会看到，它在概率论中也占有突出的地位.

下面指出一些重要分布的特征函数.

[例 1] 退化分布 $I(x-c)$ 的特征函数为

$$f(t) = e^{ict} \tag{4.5.5}$$

[例 2] 二项分布 $B(n, p)$ 的特征函数为

$$f(t) = (pe^{it} + q)^n \tag{4.5.6}$$

[例 3]　泊松分布 $P(\lambda)$ 的特征函数为

$$f(t) = e^{\lambda(e^{it}-1)} \tag{4.5.7}$$

[例 4]　Γ 分布 $G(\lambda, r)$ 的特征函数为

$$\begin{aligned} f(t) &= \int_0^{\infty} e^{itx} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} dx \\ &= \int_0^{\infty} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda(1-\frac{it}{\lambda})x} dx \\ &= \left(1 - \frac{it}{\lambda}\right)^{-r} \end{aligned} \tag{4.5.8}$$

二、性质

下面介绍特征函数的一些基本性质.

性质 1　特征函数 $f(t)$ 有如下性质:

$$f(0) = 1 \tag{4.5.9}$$

$$|f(t)| \leqslant f(0) \tag{4.5.10}$$

$$f(-t) = \overline{f(t)} \tag{4.5.11}$$

[证明]　$f(0) = \int_{-\infty}^{\infty} 1 dF(x) = 1$

$$|f(t)| \leqslant \int_{-\infty}^{\infty} |e^{itx}| dF(x) = 1 = f(0)$$

$$\begin{aligned} f(-t) &= \int_{-\infty}^{\infty} e^{-itx} |dF(x) \\ &= \overline{\int_{-\infty}^{\infty} e^{itx} dF(x)} = \overline{f(t)} \end{aligned}$$

性质 2　特征函数在 $(-\infty, \infty)$ 上一致连续.

[证明]　因为

$$|f(t+h) - f(t)| = \left| \int_{-\infty}^{\infty} (e^{i(t+h)x} - e^{itx}) dF(x) \right|$$

$$\leqslant \int_{-\infty}^{\infty} |e^{ihx} - 1| dF(x) \leqslant 2 \int_{|x| \geqslant A} dF(x) + \int_{-A}^{A} |e^{ihx} - 1| dF(x)$$

$$=2\int_{|x|\geqslant A}\mathrm{d}F(x)+2\int_{-A}^{A}\left|\sin\frac{hx}{2}\right|\mathrm{d}F(x)$$

注意上式右边已与 t 无关；可选足够大的 A 使 $\int_{|x|\geqslant A}\mathrm{d}F(x)$ 任意小，然后选充分小的 $|h|$ 可使第二个积分也任意小，从而证明了结论.

性质 3　对于任意的正整数 n 及任意实数 $t_1, t_2, \cdots, t_n$ 及复数 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 成立

$$\sum_{k=1}^{n}\sum_{j=1}^{n}f(t_k-t_j)\lambda_k\bar{\lambda}_j\geqslant 0 \tag{4.5.12}$$

［证明］

$$\sum_{k=1}^{n}\sum_{j=1}^{n}f(t_k-t_j)\lambda_k\bar{\lambda}_j$$

$$=\sum_{k=1}^{n}\sum_{j=1}^{n}\left\{\int_{-\infty}^{\infty}\mathrm{e}^{\mathrm{i}(t_k-t_j)x}\mathrm{d}F(x)\right\}\lambda_k\bar{\lambda}_j$$

$$=\int_{-\infty}^{\infty}\left\{\sum_{k=1}^{n}\sum_{j=1}^{n}\mathrm{e}^{\mathrm{i}(t_k-t_j)x}\lambda_k\bar{\lambda}_j\right\}\mathrm{d}F(x)$$

$$=\int_{-\infty}^{\infty}\left(\sum_{k=1}^{n}\mathrm{e}^{\mathrm{i}t_kx}\lambda_k\right)\left(\sum_{j=1}^{n}\mathrm{e}^{-\mathrm{i}t_jx}\bar{\lambda}_j\right)\mathrm{d}F(x)$$

$$=\int_{-\infty}^{\infty}\left|\sum_{k=1}^{n}\mathrm{e}^{\mathrm{i}t_kx}\lambda_k\right|^2\mathrm{d}F(x)\geqslant 0$$

这个性质称为**非负定性**，以后我们将会看到，这是特征函数最本质的性质之一.

性质 4　两个相互独立的随机变量之和的特征函数等于它们的特征函数之积.

［证明］　设 ξ_1 与 ξ_2 是两个相互独立的随机变量，而 $\eta=\xi_1+\xi_2$，由 ξ_1 与 ξ_2 的独立性不难推得复随机变量 $\mathrm{e}^{\mathrm{i}t\xi_1}$ 与 $\mathrm{e}^{\mathrm{i}t\xi_2}$ 也是独立的，因此

$$E\mathrm{e}^{\mathrm{i}t\eta}=E\mathrm{e}^{\mathrm{i}t(\xi_1+\xi_2)}=E\mathrm{e}^{\mathrm{i}t\xi_1}\cdot E\mathrm{e}^{\mathrm{i}t\xi_2}$$

性质 4 可推广到 n 个独立随机变量之和的场合.

应当着重指出，正是由于性质 4，才使特征函数在概率论中占有重要地位. 由于这个性质，独立随机变量和的特征函数可以方便地用各个特征函数相乘来求得，而独立和的分布函数要通过褶积这种复杂的运算才能得到，相比之下，用特征函数来处理独立和问题就有力得多. 独立和问题在概率论的古典问题中占有“中心”地位，而这些问题的解决大大有赖于特征函数的引进.

性质 5 设随机变量 ξ 有 n 阶矩存在，则它的特征函数可微分 n 次，且当 $k\leqslant n$ 时：

$$f^{(k)}(0) = \mathrm{i}^k E\xi^k \tag{4.5.13}$$

［证明］

$$\left|\frac{\mathrm{d}^k}{\mathrm{d}t^k}(\mathrm{e}^{\mathrm{i}tx})\right| = |\mathrm{i}^k x^k \mathrm{e}^{\mathrm{i}tx}| \leqslant |x|^k$$

由于 ξ 的 k 阶矩存在，故 $\int_{-\infty}^{\infty}|x|^k \mathrm{d}F(x)<\infty$，因而可作下列积分号下的微分

$$f^{(k)}(t)=\int_{-\infty}^{\infty}\frac{\mathrm{d}^k}{\mathrm{d}t^k}(\mathrm{e}^{\mathrm{i}tx})\mathrm{d}F(x)$$

$$=\mathrm{i}^k\int_{-\infty}^{\infty}x^k\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x)$$

取 $t=0$ 即得（4.5.13）.

性质 5 使我们可以方便地求得随机变量的各阶矩.

推论 若随机变量 ξ 有 n 阶矩存在，则它的特征函数可作如下展开：

$$f(t) = 1 + (\mathrm{i}t)E\xi + \frac{(\mathrm{i}t)^2}{2!}E\xi^2 + \cdots + \frac{(\mathrm{i}t)^n}{n!}E\xi^n + o(t) \tag{4.5.14}$$

［证明］ 由性质 5，$f(t)$ 可以在 $t=0$ 近旁作泰勒展开，公式（4.5.14）就是在带有皮阿诺余项的展开式中，代入（4.5.13）式而得到的.

性质 6 设 $\eta=a\xi+b$，这里 a，b 为常数，则

$$f_\eta(t) = e^{ibt} f_\xi(at) \tag{4.5.15}$$

[证明]

$$f_\eta(t) = Ee^{it\eta} = Ee^{it(a\xi+b)}$$
$$= e^{itb} Ee^{ita\xi} = e^{ibt} f_\xi(at)$$

[例 5]　正态分布 $N(a, \sigma^2)$ 的特征函数

先讨论 $N(0, 1)$ 的场合：

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{itx} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \cos tx \cdot e^{-\frac{x^2}{2}} dx$$

由于正态分布一阶矩存在，可对上式求导，得

$$f'(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-x) \sin tx e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \sin tx d e^{-\frac{x^2}{2}}$$

$$= \left[\frac{1}{\sqrt{2\pi}} \sin tx \cdot e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t \cos tx \cdot e^{-\frac{x^2}{2}} dx$$

$$= -tf(t)$$

因此

$$\ln f(t) = -\frac{t^2}{2} + c$$

由于 $f(0) = 1$，所以 $c=0$. 这样一来，

$$f(t) = e^{-\frac{t^2}{2}} \tag{4.5.16}$$

一般 $N(a, \sigma^2)$ 的场合，利用性质 6 即得

$$f(t) = e^{iat - \frac{1}{2}\sigma^2 t^2} \tag{4.5.17}$$

三、逆转公式与唯一性定理

现在来证明特征函数与分布函数是相互唯一确定的，由分布函数决定特征函数是显然的，剩下来的是需要证明可由特征函数唯一决定分布函数.

下面定理的证明要用到如下数学分析的引理.

引理 4.5.1　设 $x_1 < x_2$,

$$g(T,x,x_1,x_2)=\frac{1}{\pi}\int_0^T\left[\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right]\mathrm{d}t \tag{4.5.18}$$

则

$$\lim_{T\to\infty}g(T,x,x_1,x_2)=\begin{cases}0, & x<x_1 \text{ 或 } x>x_2\\ \frac{1}{2}, & x=x_1 \text{ 或 } x=x_2\\ 1, & x_1<x<x_2\end{cases} \tag{4.5.19}$$

［证明］　从数学分析中知道狄利克雷积分

$$D(\alpha)=\frac{1}{\pi}\int_0^{\infty}\frac{\sin\alpha t}{t}\mathrm{d}t=\begin{cases}\frac{1}{2}, & \alpha>0\\ 0, & \alpha=0\\ -\frac{1}{2}, & \alpha<0\end{cases} \tag{4.5.20}$$

而

$$\lim_{T\to\infty}g(T,x,x_1,x_2)=D(x-x_1)-D(x-x_2)$$

分别考察 x 在区间 (x_1,x_2) 的端点及内外时相应狄利克雷积分的值即得 (4.5.19).

定理 4.5.1（逆转公式）　设分布函数 $F(x)$的特征函数为 $f(t)$，又 x_1，x_2 是 $F(x)$的连续点，则

$$F(x_2)-F(x_1)=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}tx_1}-\mathrm{e}^{-\mathrm{i}tx_2}}{\mathrm{i}t}f(t)\mathrm{d}t \tag{4.5.21}$$

［证明］　不妨设 $x_1<x_2$，记

$$\begin{aligned}I_T&=\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}tx_1}-\mathrm{e}^{-\mathrm{i}tx_2}}{\mathrm{i}t}f(t)\mathrm{d}t\\&=\frac{1}{2\pi}\int_{-T}^{T}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\mathrm{i}tx_1}-\mathrm{e}^{-\mathrm{i}tx_2}}{\mathrm{i}t}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F(x)\mathrm{d}t\end{aligned}$$

为证被积函数的有界性，用到不等式

$$|e^{i\alpha}-1|\leqslant|\alpha|$$

事实上，对 $\alpha>0$

$$|e^{i\alpha}-1|=\left|\int_0^\alpha e^{ix}dx\right|\leqslant\int_0^\alpha|e^{ix}|dx=\alpha$$

对 $\alpha\leqslant 0$，取共轭即知也成立.

因此

$$\left|\frac{e^{-itx_1}-e^{-itx_2}}{it}e^{itx}\right|\leqslant x_2-x_1$$

交换上述二次积分顺序得到

$$\begin{aligned}I_T&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_{-T}^{T}\frac{e^{-itx_1}-e^{-itx_2}}{it}e^{itx}dt\right]dF(x)\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\int_0^T\frac{e^{it(x-x_1)}-e^{-it(x-x_1)}-e^{it(x-x_2)}+e^{-it(x-x_2)}}{it}dt\right]dF(x)\\&=\frac{1}{\pi}\int_{-\infty}^{\infty}\left[\int_0^T\left(\frac{\sin t(x-x_1)}{t}-\frac{\sin t(x-x_2)}{t}\right)dt\right]dF(x)\\&=\int_{-\infty}^{\infty}g(T,x,x_1,x_2)dF(x)\end{aligned}$$

此处 $g(T, x, x_1, x_2)$ 按（4.5.18）定义. 由（4.5.19）知 $|g(T, x, x_1, x_2)|$ 有界，因此由勒贝格控制收敛定理① 并利用引理的结果可得：

$$\begin{aligned}\lim_{T\to\infty}I_T&=\int_{-\infty}^{\infty}\lim_{T\to\infty}g(T,x,x_1,x_2)dF(x)\\&=F(x_2)-F(x_1)\end{aligned}$$

定理 4.5.2（唯一性定理） 分布函数由其特征函数唯一决定.

［证明］ 应用逆转公式，在 $F(x)$ 的每一连续点上，当 y 沿 $F(x)$ 的连续点趋于 $-\infty$ 时，有

$$F(x)=\frac{1}{2\pi}\lim_{y\to-\infty}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ity}-e^{-itx}}{it}f(t)dt \quad (4.5.22)$$

而分布函数由其连续点上的值唯一决定.

---

① 为了避免冗长的分析论证，在证明特征函数的逆转公式及逆极限定理时，我们共四次使用了实变函数论中关于极限号与积分号交换的勒贝格控制收敛定理.

由唯一性定理可知特征函数也完整地描述了随机变量.

特别当 $f(t)$是绝对可积函数时，有下列更强的结果.

定理 4.5.3 若$\int_{-\infty}^{\infty}|f(t)|\mathrm{d}t<\infty$,则相应的分布函数$F(x)$的导数存在并连续，而且

$$F'(x)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}tx}f(t)\mathrm{d}t \qquad (4.5.23)$$

［证明］ 由逆转公式，若 $x+\Delta x$ 及 x 是 $F(x)$的连续点，则

$$\frac{F(x+\Delta x)-F(x)}{\Delta x}=\lim_{T\to\infty}\frac{1}{2\pi}\int_{-T}^{T}\frac{\mathrm{e}^{-\mathrm{i}tx}-\mathrm{e}^{-\mathrm{i}t(x+\Delta x)}}{\mathrm{i}t\Delta x}f(t)\mathrm{d}t$$

利用$|e^{\mathrm{i}\alpha}-1|\leqslant|\alpha|$,可得

$$\left|\frac{\mathrm{e}^{-\mathrm{i}tx}-\mathrm{e}^{-\mathrm{i}t(x+\Delta x)}}{\mathrm{i}t\Delta x}\right|\leqslant 1.$$

依假设$\int_{-\infty}^{\infty}|f(t)|\mathrm{d}t<\infty$,因此

$$\frac{F(x+\Delta x)-F(x)}{\Delta x}=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\mathrm{e}^{-\mathrm{i}tx}-\mathrm{e}^{-\mathrm{i}t(x+\Delta x)}}{\mathrm{i}t\Delta x}f(t)\mathrm{d}t$$

利用控制收敛定理

$$\begin{aligned}F'(x)&=\lim_{\Delta x\to 0}\frac{F(x+\Delta x)-F(x)}{\Delta x}\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\lim_{\Delta x\to 0}\frac{\mathrm{e}^{-\mathrm{i}tx}-\mathrm{e}^{-\mathrm{i}t(x+\Delta x)}}{\mathrm{i}t\Delta x}f(t)\mathrm{d}t\\&=\frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{e}^{-\mathrm{i}tx}f(t)\mathrm{d}t\end{aligned}$$

因此 $p(x)=F'(x)$存在而有界. 再次利用控制收敛定理即得 $F'(x)$的连续性.

因此在 $f(t)$是绝对可积的条件下，分布密度 $p(x)$与特征函数 $f(t)$通过傅里叶变换来联系.

四、分布函数的再生性

许多重要的分布函数具有一个有趣的性质——再生性. 这个性质用特征函数来研究最为方便. 下面通过几个例子来说明它.

［例 6］ 若 ξ_1 服从 $B(m, p)$，ξ_2 服从 $B(n, p)$，而且 ξ_1 与 ξ_2 独立，则 $\eta=\xi_1+\xi_2$ 服从 $B(m+n, p)$.

事实上 $f_{\xi_1}(t)=(pe^{it}+q)^m, f_{\xi_2}(t)=(pe^{it}+q)^n$，由性质 4 知

$$f_\eta(t)=(pe^{it}+q)^{m+n}$$

因此由唯一性定理知 η 服从 $B(m+n,p)$.

这个事实简记作

$$B(n_1,p) * B(n_2,p) = B(n_1+n_2,p) \qquad (4.5.24)$$

［例 7］ 若 ξ_1 服从 $P(\lambda_1)$，ξ_2 服从 $P(\lambda_2)$，而且 ξ_1 与 ξ_2 独立，则 $\eta=\xi_1+\xi_2$ 服从 $P(\lambda_1+\lambda_2)$. 事实上

$$f_{\xi_1}(t)=e^{\lambda_1(e^{it}-1)}, f_{\xi_2}(t)=e^{\lambda_2(e^{it}-1)}$$

$$f_\eta(t)=e^{(\lambda_1+\lambda_2)(e^{it}-1)}$$

这个事实简记作

$$P(\lambda_1) * P(\lambda_2) = P(\lambda_1+\lambda_2) \qquad (4.5.25)$$

［例 8］ 若 ξ_1 服从 $N(a_1, \sigma_1^2)$，ξ_2 服从 $N(a_2, \sigma_2^2)$，而且 ξ_1 与 ξ_2 独立，则 $\eta=\xi_1+\xi_2$ 服从 $N(a_1+a_2, \sigma_1^2+\sigma_2^2)$.

事实上

$$f_{\xi_1}(t)=e^{ia_1t-\frac{1}{2}\sigma_1^2t^2}, f_{\xi_2}(t)=e^{ia_2t-\frac{1}{2}\sigma_2^2t^2}$$

$$f_\eta(t)=e^{i(a_1+a_2)t-\frac{1}{2}(\sigma_1^2+\sigma_2^2)t^2}$$

这个事实简记作

$$N(a_1,\sigma_1^2) * N(a_2,\sigma_2^2)=N(a_1+a_2,\sigma_1^2+\sigma_2^2) \qquad (4.5.26)$$

［例 9］ 若 ξ_1 服从 $G(\lambda,r_1)$，ξ_2 服从 $G(\lambda, r_2)$，而且 ξ_1 与 ξ_2 独立，则 $\eta=\xi_1+\xi_2$ 服从 $G(\lambda,r_1+r_2)$.

事实上，由 (4.5.8)

$$f_{\xi_1}(t)=\left(1-\frac{it}{\lambda}\right)^{-r_1}, f_{\xi_2}(t)=\left(1-\frac{it}{\lambda}\right)^{-r_2}$$

$$f_\eta(t)=\left(1-\frac{it}{\lambda}\right)^{-(r_1+r_2)}$$

这个事实简记作

$$G(\lambda_1, r_1) * G(\lambda, r_2) = G(\lambda, r_1 + r_2) \qquad (4.5.27)$$

还有不少重要分布也有再生性，我们把它们留给读者作为练习.

还有人研究了这类命题的逆命题——分布函数的分解问题，即若两个独立随机变量之和服从某一分布，问是否能断定这两个随机变量也分别服从这个分布. 已经证明对于正态分布及泊松分布逆命题的确成立.

五、多元特征函数

若随机向量$(\xi_1, \xi_2, \cdots, \xi_n)$的分布函数为$F(x_1, x_2, \cdots, x_n)$，与随机变量相仿，我们可以定义它的特征函数

$$f(t_1, t_2, \cdots, t_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{i(t_1 x_1 + \cdots + t_n x_n)} dF(x_1, \cdots, x_n) \qquad (4.5.28)$$

可以类似于一元的场合，建立起 n 元特征函数的理论，由于方法完全相同，我们只叙述一些有关结论，证明一概从略.

性质 1 $f(t_1, t_2, \cdots, t_n)$在 $\mathbf{R}^n$ 中一致连续，而且

$$|f(t_1, t_2, \cdots, t_n)| \leqslant f(0, 0, \cdots, 0) = 1$$

$$f(-t_1, -t_2, \cdots, -t_n) = \overline{f(t_1, t_2, \cdots, t_n)}$$

性质 2 如果 $f(t_1, t_2, \cdots, t_n)$是$(\xi_1, \xi_2, \cdots, \xi_n)$的特征函数，则 $\eta = a_1\xi_1 + a_2\xi_2 + \cdots + a_n\xi_n$ 的特征函数为

$$f_\eta(t) = f(a_1 t, a_2 t, \cdots, a_n t)$$

性质 3 如果矩 $E\xi_1^{k_1}\xi_2^{k_2}\cdots\xi_n^{k_n}$ 存在，则

$$E\xi_1^{k_1}\xi_2^{k_2}\cdots\xi_n^{k_n} = i^{-\sum_{j=1}^{n} k_j} \left[\frac{\partial^{k_1 + k_2 + \cdots + k_n} f(t_1, t_2, \cdots, t_n)}{\partial t_1^{k_1} \partial t_2^{k_2} \cdots \partial t_n^{k_n}} \right]_{t_1 = t_2 = \cdots = t_n = 0} \qquad (4.5.29)$$

性质 4 若 $(\xi_1, \xi_2, \cdots, \xi_n)$ 的特征函数为 $f(t_1, t_2, \cdots, t_n)$，则 k $(k<n)$ 维随机向量 $(\xi_1, \xi_2, \cdots, \xi_k)$ 的特征函数为

$$f_{1,2,\cdots,k}(t_1, t_2, \cdots, t_k) = f(t_1, t_2, \cdots, t_k, 0, \cdots, 0)$$

这是前 k 个分量的 k 元边际分布函数对应的特征函数. 对应于任意 k 个分量 ξ_{j_1}，ξ_{j_2}，…，ξ_{j_k}的边际分布函数的特征函数，可以类似得到.

逆转公式 如果 $f(t_1, t_2, \cdots, t_n)$ 是随机向量（ξ_1，ξ_2，…，ξ_n）的特征函数，而 $F(x_1, x_2, \cdots, x_n)$ 是它的分布函数，则

$$P\{a_k \leqslant \xi_k < b_k, k=1,2,\cdots,n\}$$
$$= \lim_{\substack{T_j \to \infty \\ j=1,\cdots,n}} \frac{1}{(2\pi)^n} \int_{-T_1}^{T_1} \int_{-T_2}^{T_2} \cdots \int_{-T_n}^{T_n} \prod_{k=1}^{n} \frac{e^{-it_k a_k} - e^{-it_k b_k}}{it_k}$$
$$\cdot f(t_1, t_2, \cdots, t_n) dt_1 dt_2 \cdots dt_n$$

其中 a_k 和 b_k 都是任意实数，但满足唯一的要求：（ξ_1，ξ_2，…，ξ_n）落在平行体

$$a_k \leqslant x_k < b_k, \quad k=1, 2, \cdots, n$$

的面上的概率等于零.

唯一性定理 分布函数 $F(x_1, x_2, \cdots, x_n)$由其特征函数唯一决定.

有了唯一性定理，可以进一步证明特征函数的如下两个性质，它们表征了独立性.

性质 5 若(ξ_1，ξ_2，…，ξ_n) 的特征函数为 $f(t_1, t_2, \cdots, t_n)$，而 ξ_j 的特征函数为 $f_{\xi_j}(t)$，$j=1, 2, \cdots, n$，则随机变量 ξ_1，ξ_2，…，ξ_n 相互独立的充要条件为

$$f(t_1, t_2, \cdots, t_n) = f_{\xi_1}(t_1) f_{\xi_2}(t_2) \cdots, f_{\xi_n}(t_n) \qquad (4.5.30)$$

性质 6 若以 $f_1(t_1, \cdots, t_n)$，$f_2(u_1, \cdots, u_m)$ 及 $f(t_1, \cdots, t_n, u_1, \cdots, u_m)$ 分别记随机向量（ξ_1，…，ξ_n），（η_1，…，η_m）及（ξ_1，…，ξ_n，η_1，…，η_m）的特征函数，则（ξ_1，…，ξ_n）与（η_1，…，η_m）独立的充要条件为：对一切实数 t_1，…，t_n 及 u_1，…，u_m 成立

$$f(t_1, \cdots, t_n, u_1, \cdots, u_m)$$
$$= f_1(t_1, \cdots, t_n) f_2(u_1, \cdots, u_m) \qquad (4.5.31)$$

在下一节，我们要用到如下定理，相应于此定理的一维结果，将在下章叙述并证明.

连续性定理　若特征函数列 $\{f_k(t_1, t_2, \cdots, t_n)\}$ 收敛于一个连续函数 $f(t_1, t_2, \cdots, t_n)$，则函数 $f(t_1, t_2, \cdots, t_n)$ 是某分布函数所对应的特征函数.

*§6. 多元正态分布

一、密度函数与特征函数

在本节中，我们将讨论多元正态分布的定义与性质，假定读者具有矩阵论的基本知识. 因为对二元的场合我们已直接推导过其中的大部分结果，所以基础知识不足的读者不妨跳过本节.

下面，我们将以黑体的小写字母记列向量

$$\boldsymbol{a}=\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}, \quad \boldsymbol{\xi}=\begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{pmatrix}$$

以黑体的大写字母记矩阵，例如

$$\boldsymbol{B}=\begin{pmatrix} b_{11} & b_{12}\cdots b_{1n} \\ b_{21} & b_{22}\cdots b_{2n} \\ \cdots\cdots & \cdots\cdots \\ b_{n1} & b_{n2}\cdots b_{nn} \end{pmatrix}$$

以 $'$ 记矩阵（当然也包括向量）的转置，因此 $\boldsymbol{t}'\boldsymbol{\xi}=\sum_{i=1}^{n} t_i\xi_i$（其中，$\boldsymbol{t}$ 为列向量）；以 $\boldsymbol{B}^{-1}$ 记 $\boldsymbol{B}$ 的逆阵；以 $\det\boldsymbol{B}$ 记 $\boldsymbol{B}$ 的行列式之值.

在 (3.2.13) 中，我们已定义了 n 元正态分布的密度函数，采用列向量形式，其表达式为

$$p(\boldsymbol{x})=\frac{1}{(2\pi)^{n/2}(\det\boldsymbol{B})^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})\right\} \tag{4.6.1}$$

其中 $\boldsymbol{B}$ 是 n 阶正定对称矩阵，$\boldsymbol{a}$ 是实值列向量，并简记为 $N(\boldsymbol{a},\boldsymbol{B})$.

事实上，我们还需要证明由（4.6.1）定义的函数是 R^n 中的密度函数.

显然

$$p(\boldsymbol{x})>0,\quad \boldsymbol{x}\in \mathbf{R}^n \tag{4.6.2}$$

因此只须验证下式成立

$$\int_{R^n} p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}=1 \tag{4.6.3}$$

为了证明（4.6.3），我们要用到矩阵论中这样一个结果：若 $\boldsymbol{B}$ 是正定对称阵，则存在非奇异阵 $\boldsymbol{L}$，使

$$\boldsymbol{B}=\boldsymbol{L}\boldsymbol{L}' \tag{4.6.4}$$

作线性变换

$$\boldsymbol{y}=\boldsymbol{L}^{-1}(\boldsymbol{x}-\boldsymbol{a}) \tag{4.6.5}$$

则逆变换为

$$\boldsymbol{x}=\boldsymbol{L}\boldsymbol{y}+\boldsymbol{a} \tag{4.6.6}$$

变换(4.6.6)的雅可比行列式由下式给出

$$\det\boldsymbol{L}=(\det\boldsymbol{B})^{\frac{1}{2}} \tag{4.6.7}$$

因此

$$\begin{aligned}\int_{R^n} p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} &= \frac{1}{(2\pi)^{n/2}(\det\boldsymbol{B})^{\frac{1}{2}}}\int_{R^n}\exp\left\{-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}\right\}\cdot(\det\boldsymbol{B})^{\frac{1}{2}}\mathrm{d}\boldsymbol{y}\\ &=\frac{1}{(2\pi)^{n/2}}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n}y_k^2\right\}\mathrm{d}y_1\cdots\mathrm{d}y_n\\ &=\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-u^2/2}\mathrm{d}u\right]^n=1\end{aligned}$$

从而证明了（4.6.3），所以（4.6.1）确实定义了 $\mathbf{R}^n$ 上的一个密度函数.

定理 4.6.1 n 元正态分布（4.6.1）的特征函数为

$$f(\boldsymbol{t})=\exp\left\{\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}\right\} \tag{4.6.8}$$

［证明］ 按定义

$$f(\boldsymbol{t})=\int_{R^n}\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{x}}p(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$

$$=\frac{1}{(2\pi)^{n/2}\ (\det\boldsymbol{B})^{\frac{1}{2}}}$$

$$\cdot\int_{R^n}\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{x}}\exp\left\{-\frac{1}{2}\ (\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{B}^{-1}\ (\boldsymbol{x}-\boldsymbol{a})\right\}\mathrm{d}\boldsymbol{x}$$

作变换（4.6.5），注意到

$$\mathrm{i}\boldsymbol{t}'\boldsymbol{x}-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})$$

$$=\mathrm{i}\boldsymbol{t}'\boldsymbol{a}+\mathrm{i}\boldsymbol{t}'\boldsymbol{L}\boldsymbol{y}-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}$$

记 $\boldsymbol{s}=\boldsymbol{L}'\boldsymbol{t}$，则

$$\mathrm{i}\boldsymbol{t}'\boldsymbol{x}-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{a})'\boldsymbol{B}^{-1}(\boldsymbol{x}-\boldsymbol{a})$$

$$=\mathrm{i}\boldsymbol{t}'\boldsymbol{a}+\mathrm{i}\boldsymbol{s}'\boldsymbol{y}-\frac{1}{2}\boldsymbol{y}'\boldsymbol{y}$$

$$=\mathrm{i}\boldsymbol{a}'\boldsymbol{t}+\sum_{k=1}^{n}\left(\mathrm{i}s_k y_k-\frac{1}{2}y_k^2\right)$$

$$=\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\sum_{k=1}^{n}(y_k-\mathrm{i}s_k)^2-\frac{1}{2}\sum_{k=1}^{n}s_k^2$$

$$=\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\sum_{k=1}^{n}(y_k-\mathrm{i}s_k)^2-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}$$

因此

$$f(\boldsymbol{t})=\frac{\mathrm{e}^{\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}}}{(2\pi)^{n/2}(\det\boldsymbol{B})^{\frac{1}{2}}}$$

$$\cdot\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\sum_{k=1}^{n}(y_k-\mathrm{i}s_k)^2\right\}(\det\boldsymbol{B})^{\frac{1}{2}}\mathrm{d}y_1\cdots\mathrm{d}y_n$$

$$=\mathrm{e}^{\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}}$$

在（4.6.1）中，假定 $\boldsymbol{B}$ 是正定对称阵，否则该表达式没有意义，因此我们只是对正定对称阵 $\boldsymbol{B}$ 的场合定义了多元正态分布，但是利用特征函数表达式（4.6.8）有可能把定义拓广到一般非负

定对称阵 $\boldsymbol{B}$ 的场合.

事实上，设 $\boldsymbol{B}$ 是非负定对称阵，令 $\boldsymbol{B}_k=\boldsymbol{B}+\frac{1}{k}\boldsymbol{I}$，这里 $\boldsymbol{I}$ 是 n 阶单位阵，显然 $\boldsymbol{B}_k$ 是正定对称阵，因此

$$f_k(\boldsymbol{t})=\mathrm{e}^{\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}_k\boldsymbol{t}}$$

是 n 元正态分布 $N(\boldsymbol{a},\boldsymbol{B}_k)$的特征函数. 现在

$$\lim_{k\to\infty}f_k(\boldsymbol{t})=\mathrm{e}^{\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}}=f(\boldsymbol{t})$$

而 $f(\boldsymbol{t})$在 R^n 上连续,因此由上节的连续性定理知 $f(\boldsymbol{t})$是 R^n 上某分布函数的特征函数.

这样可以引进如下一般定义.

定义 4.6.1 若 $\boldsymbol{a}$ 是 n 维实向量，$\boldsymbol{B}$ 是 n 阶非负定对称阵，则称以（4.6.8）式中的 f ($\boldsymbol{t}$) 为其特征函数的分布函数为 **n 元正态分布**，并简记为 $N(\boldsymbol{a},\boldsymbol{B})$.

按照这个定义，当 $\boldsymbol{B}$ 为正定对称阵时，其密度函数由(4.6.1)给出；但是当 $\det\boldsymbol{B}=0$ 时，密度函数无法写出. 可以证明，若 $\boldsymbol{B}$ 的秩为 r ($r<n$)，则这时概率分布集中在一个 r 维子空间上，这种正态分布称为**退化正态分布**或**奇异正态分布**.

在下面讨论中，我们总是假定随机向量 $\boldsymbol{\xi}=(\xi_1, \xi_2, \cdots, \xi_n)'$服从 n 元正态分布 N ($\boldsymbol{a}$, $\boldsymbol{B}$).

定理 4.6.2 $\boldsymbol{\xi}$ 的任一子向量 $(\xi_{k_1}, \xi_{k_2}, \cdots, \xi_{k_m})'$ ($m\leqslant n$) 也服从正态分布，分布为 N ($\tilde{\boldsymbol{a}}$, $\tilde{\boldsymbol{B}}$)，其中 $\tilde{\boldsymbol{a}}=(a_{k_1}, a_{k_2}, \cdots, a_{k_m})'$，$\tilde{\boldsymbol{B}}$ 为保留 $\boldsymbol{B}$ 的第 k_1, k_2, $\cdots$, k_m 行及列所得的 m 阶矩阵.

特别地，ξ_j 服从一元正态分布 N (a_j, b_{jj}).

[证明] 只须在特征函数（4.6.8）中对一切不等于 k_1, k_2, $\cdots$, k_m 的 l，令 $t_l=0$ 即得 $(\xi_{k_1}, \xi_{k_2}, \cdots, \xi_{k_m})'$的特征函数

$$\tilde{f}(\tilde{\boldsymbol{t}})=\mathrm{e}^{\mathrm{i}\tilde{\boldsymbol{a}}'\tilde{\boldsymbol{t}}-\frac{1}{2}\tilde{\boldsymbol{t}}'\tilde{\boldsymbol{B}}\tilde{\boldsymbol{t}}}$$

这里 $\tilde{\boldsymbol{t}}=(t_{k_1},t_{k_2},\cdots,t_{k_m})'$,这正是 $N(\tilde{\boldsymbol{a}},\tilde{\boldsymbol{B}})$的特征函数.

定理 4.6.2 表明,多元正态分布的边际分布还是正态分布.

定理 4.6.3 $\boldsymbol{a}$ 及 $\boldsymbol{B}$ 分别是随机向量 $\boldsymbol{\xi}$ 的数学期望及协方差矩阵,即

$$a_j=E\xi_j,\qquad 1\leqslant j\leqslant n \tag{4.6.9}$$

$$b_{jk}=E(\xi_j-a_j)(\xi_k-a_k),\quad 1\leqslant j,k\leqslant n \tag{4.6.10}$$

[证明] 由定理 4.6.2 立即得到 (4.6.9),而且知道 $D\xi_j=b_{jj}$ 存在,由柯西—施瓦兹不等式又知各协方差存在,因此由 (4.5.29)

$$E\xi_j\xi_k=\frac{1}{\mathrm{i}^2}\cdot\left.\frac{\partial^2 f(t_1,\cdots,t_n)}{\partial t_j\partial t_k}\right|_{t_1=\cdots=t_n=0}=b_{jk}+a_ja_k$$

故

$$E(\xi_j-a_j)(\xi_k-a_k)=E\xi_j\xi_k-a_ja_k=b_{jk}$$

因此 n 元正态分布由它的前面二阶矩完全确定.

二、独立性

正如在二元场合一样,对 n 元正态分布而言,独立性与不相关性有密切联系.

定理 4.6.4 ξ_1, ξ_2, …, ξ_n 相互独立的充要条件是它们两两不相关.

[证明] 必要性显然成立,下证充分性.

若 ξ_1, ξ_2, …, ξ_n 两两不相关,即对一切 $j\neq k$,

$$r_{jk}=\frac{E[(\xi_j-E\xi_j)(\xi_k-E\xi_k)]}{\sqrt{D\xi_j}\sqrt{D\xi_k}}=0$$

因此 $b_{jk}=E(\xi_j-E\xi_j)(\xi_k-E\xi_k)=0$,所以

$$\begin{aligned}f(t_1,\cdots,t_n)&=\mathrm{e}^{\mathrm{i}\sum\limits_{k=1}^{n}a_kt_k-\frac{1}{2}\sum\limits_{k=1}^{n}b_{kk}t_k^2}\\&=\prod_{k=1}^{n}\mathrm{e}^{\mathrm{i}a_kt_k-\frac{1}{2}b_{kk}t_k^2}\\&=\prod_{k=1}^{n}f_{\xi_k}(t_k)\end{aligned}$$

由上节多元特征函数的性质 5 可知 ξ_1, ξ_2, …, ξ_n 相互独立.

定理 4.6.5 若$\xi=\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$，这里$\xi_1$与$\xi_2$是$\xi$的子向量，记

$$\boldsymbol{B}=\begin{pmatrix}\boldsymbol{B}_{11} & \boldsymbol{B}_{12}\\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22}\end{pmatrix} \tag{4.6.11}$$

其中$\boldsymbol{B}_{11}$及$\boldsymbol{B}_{22}$分别是ξ_1及ξ_2的协方差矩阵，$\boldsymbol{B}_{12}$则是由ξ_1与ξ_2的相应分量的协方差构成的相互协方差矩阵，则ξ_1与ξ_2独立的充要条件是$\boldsymbol{B}_{12}=\boldsymbol{0}$.

[证明] 若ξ_1与ξ_2独立，则ξ_1的任一分量与ξ_2的任一分量独立，因此其协方差为0，从而由它们构成的矩阵$\boldsymbol{B}_{12}=\boldsymbol{0}$，这就证明了必要性.

下面来证明充分性. 由$b_{ij}=b_{ji}$，因此$\boldsymbol{B}_{21}=\boldsymbol{B}_{12}'=\boldsymbol{0}$，若$\boldsymbol{t}=\begin{pmatrix}\boldsymbol{t}_1\\ \boldsymbol{t}_2\end{pmatrix}$，这里$\boldsymbol{t}_1$与$\xi_1$有相同维数，$\boldsymbol{t}_2$与$\xi_2$有相同维数，则

$$\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}=\boldsymbol{t}_1'\boldsymbol{B}_{11}\boldsymbol{t}_1+2\boldsymbol{t}_1'\boldsymbol{B}_{12}\boldsymbol{t}_2+\boldsymbol{t}_2'\boldsymbol{B}_{22}\boldsymbol{t}_2=\boldsymbol{t}_1'\boldsymbol{B}_{11}\boldsymbol{t}_1+\boldsymbol{t}_2'\boldsymbol{B}_{22}\boldsymbol{t}_2$$

因此若记$\boldsymbol{a}=\begin{pmatrix}\boldsymbol{a}_1\\ \boldsymbol{a}_2\end{pmatrix}$，其中$\boldsymbol{a}_1$，$\boldsymbol{a}_2$分别为$\xi_1$及$\xi_2$的数学期望，则

$$\begin{aligned}f_\xi(\boldsymbol{t})&=\exp\left\{\mathrm{i}\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}\right\}\\&=\exp\left\{\mathrm{i}\boldsymbol{a}_1'\boldsymbol{t}_1+\mathrm{i}\boldsymbol{a}_2'\boldsymbol{t}_2-\frac{1}{2}\boldsymbol{t}_1'\boldsymbol{B}_{11}\boldsymbol{t}_1-\frac{1}{2}\boldsymbol{t}_2'\boldsymbol{B}_{22}\boldsymbol{t}_2\right\}\\&=\exp\left\{\mathrm{i}\boldsymbol{a}_1'\boldsymbol{t}_1-\frac{1}{2}\boldsymbol{t}_1'\boldsymbol{B}_{11}\boldsymbol{t}_1\right\}\cdot\exp\left\{\mathrm{i}\boldsymbol{a}_2'\boldsymbol{t}_2-\frac{1}{2}\boldsymbol{t}_2'\boldsymbol{B}_{22}\boldsymbol{t}_2\right\}\\&=f_{\xi_1}(\boldsymbol{t}_1)f_{\xi_2}(\boldsymbol{t}_2)\end{aligned}$$

由上节多元特征函数的性质6可知ξ_1与ξ_2独立.

类似地可以证明，若ξ的子向量ξ_1，ξ_2，…，ξ_k两两独立，则它们也相互独立.

三、线性变换

服从正态分布的随机向量在线性变换下具有许多特殊的性质，这些性质有很大的理论和实用价值，下面只讨论这类性质中最基本的一些.

一般，若 $\boldsymbol{\xi}=(\xi_1, \cdots, \xi_n)'$ 是 n 维随机向量，其数学期望为 $\boldsymbol{a}$，协方差矩阵为 $\boldsymbol{B}$.

考虑 $\boldsymbol{\xi}$ 的分量的线性组合 $\zeta=\sum_{j=1}^{n} l_j\xi_j=\boldsymbol{l}'\boldsymbol{\xi}$，显然

$$E\zeta=\sum_{j=1}^{n} l_j a_j=\boldsymbol{l}'\boldsymbol{a} \tag{4.6.12}$$

$$D\zeta=\sum_{j=1}^{n}\sum_{k=1}^{n} l_j l_k b_{jk}=\boldsymbol{l}'\boldsymbol{B}\boldsymbol{l} \tag{4.6.13}$$

同样地，若 $\boldsymbol{C}=(c_{jk})$ 是 $m\times n$ 矩阵，则 m 维随机向量 $\boldsymbol{\eta}=\boldsymbol{C}\boldsymbol{\xi}$ 有

$$E\boldsymbol{\eta}=\boldsymbol{C}\boldsymbol{a} \tag{4.6.14}$$

$$D\boldsymbol{\eta}=\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}' \tag{4.6.15}$$

这里 $D\boldsymbol{\eta}$ 记 $\boldsymbol{\eta}$ 的协方差矩阵.

定理 4.6.6 $\boldsymbol{\xi}=(\xi_1, \cdots, \xi_n)'$ 服从 n 元正态分布 $N(\boldsymbol{a},\boldsymbol{B})$ 的充要条件是它的任何一个线性组合 $\zeta=\sum_{j=1}^{n} l_j\xi_j$ 服从一元正态分布 $N\left(\sum_{j=1}^{n} l_j a_j,\ \sum_{j,k=1}^{n} l_j l_k b_{jk}\right)$.

[证明] 必要性：若 $\boldsymbol{\xi}$ 服从 $N(\boldsymbol{a}, \boldsymbol{B})$，则由（4.6.8）

$$Ee^{i\boldsymbol{t}'\boldsymbol{\xi}}=\exp\left\{i\boldsymbol{a}'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'\boldsymbol{B}\boldsymbol{t}\right\}$$

取 $\boldsymbol{t}=u\boldsymbol{l}$，这里 u 是实数，则

$$Ee^{iu\zeta}=Ee^{iu\boldsymbol{l}'\boldsymbol{\xi}}=\exp\left\{iu\boldsymbol{a}'\boldsymbol{l}-\frac{1}{2}u^2\boldsymbol{l}'\boldsymbol{B}\boldsymbol{l}\right\} \tag{4.6.16}$$

对 u 是任意实数上式都成立，这就说明随机变量 ζ 服从 $N(\boldsymbol{l}'\boldsymbol{a}, \boldsymbol{l}'\boldsymbol{B}\boldsymbol{l})$.

充分性：若 $\zeta=\boldsymbol{l}'\boldsymbol{\xi}$ 服从 $N(\boldsymbol{l}'\boldsymbol{a},\boldsymbol{l}'\boldsymbol{B}\boldsymbol{l})$，仍有(4.6.16)，在该式中取 $u=1$，得

$$Ee^{i\boldsymbol{l}'\boldsymbol{\xi}}=\exp\left\{i\boldsymbol{a}'\boldsymbol{l}-\frac{1}{2}\boldsymbol{l}'\boldsymbol{B}\boldsymbol{l}\right\}$$

由于 $\boldsymbol{l}$ 的任意性，这说明 $\boldsymbol{\xi}$ 服从 $N(\boldsymbol{a},\boldsymbol{B})$.

利用定理 4.6.6 可以通过一维正态变量来研究多维正态变

量. 在有些场合这提供了很大的方便.

定理 4.6.7 若$\xi=(\xi_1, \cdots, \xi_n)'$服从$n$元正态分布$N(\boldsymbol{a}, \boldsymbol{B})$，而$C$为任意$m\times n$阵，则$\eta=\boldsymbol{C}\xi$服从$m$元正态分布$N(\boldsymbol{Ca}, \boldsymbol{CBC'})$.

［证明］ 因为对于任意m维实值列向量$\boldsymbol{t}$，

$$\begin{aligned} f_\eta(\boldsymbol{t}) &= E\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\eta}=E\mathrm{e}^{\mathrm{i}\boldsymbol{t}'\boldsymbol{C}\xi}=E\mathrm{e}^{\mathrm{i}(\boldsymbol{C}'\boldsymbol{t})'\xi} \\ &=\exp\left\{\mathrm{i}\boldsymbol{a}'(\boldsymbol{C}'\boldsymbol{t})-\frac{1}{2}(\boldsymbol{C}'\boldsymbol{t})'\boldsymbol{B}(\boldsymbol{C}'\boldsymbol{t})\right\} \\ &=\exp\left\{\mathrm{i}(\boldsymbol{Ca})'\boldsymbol{t}-\frac{1}{2}\boldsymbol{t}'(\boldsymbol{CBC}')\boldsymbol{t}\right\} \end{aligned} \tag{4.6.17}$$

按定义，η服从m元正态分布$N(\boldsymbol{Ca}, \boldsymbol{CBC'})$.

定理 4.6.7 表明正态变量在线性变换下还是正态变量，这个性质简称为**正态变量的线性变换不变性**.

推论 1 若ξ服从n元正态分布$N(\boldsymbol{a}, \boldsymbol{B})$，则存在一个正交变换$\boldsymbol{U}$,使得$\eta=\boldsymbol{U}\xi$是一个具有独立正态分布分量的随机向量,它的数学期望为$\boldsymbol{Ua}$，而它的方差分量是$\boldsymbol{B}$的特征值.

［证明］ 从矩阵论知道，对实对称矩阵$\boldsymbol{B}$，存在正交阵$\boldsymbol{U}$，使$\boldsymbol{UBU'}=\boldsymbol{D}$，其中

$$\boldsymbol{D}=\begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_n \end{pmatrix} \tag{4.6.18}$$

这里$d_1, d_2, \cdots, d_n$是$\boldsymbol{B}$的特征值. 若$\boldsymbol{B}$的秩为r，则有r个特征值不为零. 此处的$\boldsymbol{U}$是以特征向量为列构成的正交阵.

把这里的$\boldsymbol{U}$作为定理 4.6.7 中的变换矩阵，则利用该定理的结果即得推论 1.

从推论 1 可以看出，若$\boldsymbol{B}$的秩$r<n$，则正态分布退化到一个r维子空间上.

推论 1 说明，对于多维正态变量，可以进行正交变换，使其既保持正态性不变又让各分量独立，这种方法在数理统计中十分

有用.

第三章§3中的例8提供了利用坐标旋转（正交变换的一种）把二维正态变量化为独立分量的生动例子.

推论 2 在正交变换下，多维正态变量保持其独立、同方差性不变.

其证明留给读者作为习题。

四、条件分布

若 $\boldsymbol{\xi}=\begin{pmatrix}\boldsymbol{\xi}_1\\ \boldsymbol{\xi}_2\end{pmatrix}$ 服从 n 元正态分布 $N(\boldsymbol{a},\boldsymbol{B})$，这里 $\boldsymbol{\xi}_1$，$\boldsymbol{\xi}_2$ 是它的子向量，$E\boldsymbol{\xi}_1=\boldsymbol{a}_1$，$E\boldsymbol{\xi}_2=\boldsymbol{a}_2$，协方差矩阵 $\boldsymbol{B}$ 仍由（4.6.11）表出.

下面求在给定 $\boldsymbol{\xi}_1=\boldsymbol{x}_1$ 的条件下 $\boldsymbol{\xi}_2$ 的分布密度函数. 假定 $\det\boldsymbol{B}\neq 0$，即只讨论非奇异的场合，这时 $\det\boldsymbol{B}_{11}\neq 0$.

首先，我们来找一个线性变换

$$\boldsymbol{\eta}_1=\boldsymbol{\xi}_1$$

$$\boldsymbol{\eta}_2=\boldsymbol{T}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2$$

这个线性变换使 $\boldsymbol{\eta}_1=\boldsymbol{\xi}_1$，此外我们还要求它使得 $\boldsymbol{\eta}_1$ 与 $\boldsymbol{\eta}_2$ 独立. 由定理4.6.7我们知道 $\boldsymbol{\eta}=\begin{pmatrix}\boldsymbol{\eta}_1\\ \boldsymbol{\eta}_2\end{pmatrix}$ 是联合正态的，因此根据定理4.6.5为使 $\boldsymbol{\eta}_1$ 与 $\boldsymbol{\eta}_2$ 独立，只须它们的相互协方差矩阵为零阵. 但是

$$\begin{aligned}&E(\boldsymbol{\eta}_1-E\boldsymbol{\eta}_1)(\boldsymbol{\eta}_2-E\boldsymbol{\eta}_2)'\\ &=E(\boldsymbol{\xi}_1-E\boldsymbol{\xi}_1)(\boldsymbol{T}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2-\boldsymbol{T}E\boldsymbol{\xi}_1-E\boldsymbol{\xi}_2)'\\ &=\boldsymbol{B}_{11}\boldsymbol{T}'+\boldsymbol{B}_{12}\end{aligned}$$

因此应取

$$\boldsymbol{T}=-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}$$

故

$$\boldsymbol{\eta}_2=-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{\xi}_1+\boldsymbol{\xi}_2$$

所求的线性变换为

$$\begin{pmatrix}\eta_1\\ \eta_2\end{pmatrix}=\begin{pmatrix}\boldsymbol{I} & \boldsymbol{0}\\ -\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1} & \boldsymbol{I}\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} \tag{4.6.19}$$

为求 η_1 及 η_2 的分布密度函数，先计算

$$E\eta_1=E\xi_1=\boldsymbol{a}_1,\ D\eta_1=D\xi_1=\boldsymbol{B}_{11}$$

$$\begin{aligned}E\eta_2&=-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}E\xi_1+E\xi_2=\boldsymbol{a}_2-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{a}_1\\ D\eta_2&=E(\eta_2-E\eta_2)(\eta_2-E\eta_2)'\\ &=E[(\xi_2-\boldsymbol{a}_2)-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}(\xi_1-\boldsymbol{a}_1)]\\ &\quad\cdot[(\xi_2-\boldsymbol{a}_2)-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}(\xi_1-\boldsymbol{a}_1)]'\\ &=\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}+\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{11}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\\ &=\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12}\end{aligned}$$

正如上述，η_1 与 η_2 是独立的，又变换（4.6.19）的雅可比行列式等于 1，所以有

$$p_\xi(\boldsymbol{x}_1,\boldsymbol{x}_2)=p_\eta(\boldsymbol{y}_1,\boldsymbol{y}_2)=p_{\eta_1}(\boldsymbol{y}_1)p_{\eta_2}(\boldsymbol{y}_2)$$

这里 $\boldsymbol{y}_1=\boldsymbol{x}_1$，$\boldsymbol{y}_2=-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{x}_1+\boldsymbol{x}_2$，显然 $p_{\xi_1}(\boldsymbol{x}_1)=p_{\eta_1}(\boldsymbol{y}_1)$，因此在给定 $\xi_1=\boldsymbol{x}_1$ 的条件下，ξ_2 的密度函数

$$\begin{aligned}p(\boldsymbol{x}_2|\xi_1=\boldsymbol{x}_1)&=\frac{p_\xi(\boldsymbol{x}_1,\boldsymbol{x}_2)}{p_{\xi_1}(\boldsymbol{x}_1)}=\frac{p_{\eta_1}(\boldsymbol{y}_1)p_{\eta_2}(\boldsymbol{y}_2)}{p_{\eta_1}(\boldsymbol{y}_1)}\\ &=p_{\eta_2}(\boldsymbol{y}_2)=p_{\eta_2}(\boldsymbol{x}_2-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{x}_1)\end{aligned}$$

因为 η_2 服从 $N(\boldsymbol{a}_2-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{a}_1,\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})$，所以 ξ_2 对于 $\xi_1=\boldsymbol{x}_1$ 的条件分布是正态分布 $N(\boldsymbol{a}_2+\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{a}_1),\ \boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12})$，我们把这个结果总结为下列定理.

定理 4.6.8 若 $\xi=\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}$ 服从 n 元正态分布 $N(\boldsymbol{a},\boldsymbol{B})$，$E\xi_1=\boldsymbol{a}_1$，$E\xi_2=\boldsymbol{a}_2$，$\boldsymbol{B}$ 表示成（4.6.11），则在给定 $\xi_1=\boldsymbol{x}_1$ 下，ξ_2 的条件分布还是正态分布，其条件数学期望

$$E(\xi_2|\xi_1=\boldsymbol{x}_1)=\boldsymbol{a}_2+\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}(\boldsymbol{x}_1-\boldsymbol{a}_1) \tag{4.6.20}$$

其条件方差

$$\boldsymbol{B}_{22\cdot1}=\boldsymbol{B}_{22}-\boldsymbol{B}_{21}\boldsymbol{B}_{11}^{-1}\boldsymbol{B}_{12} \tag{4.6.21}$$

这里 $E(\xi_2|\xi_1=x_1)$ 称为 ξ_2 关于 ξ_1 的**回归**，注意到它是 x_1 的线性函数，又条件方差 $B_{22\cdot 1}$ 与 x_1 无关.

我们已对二元正态分布的场合进行过类似的讨论，这类问题将在数理统计中继续深入研究.

第四章小结

本章主要讨论了随机变量（或分布函数）的数字特征与特征函数，它们都是概率分布的某种表征. 这些讨论不但深化了对随机变量的认识，同时也为以后的研究作了必要的准备.

数字特征是描述随机变量特征的有效工具，它虽然不像分布函数那样完整地描述了随机变量，但是却具有很多优点：它较集中地反映了随机变量变化的一些平均特征（事实上数字特征多是某种平均值）；其次，大部分最重要的分布函数都由一、两个数字特征完全确定，而数字特征较易求得；最后，最常用也是最重要的几个数字特征——数学期望、方差、相关系数——都有明确的概率意义，同时又具有良好的性质. 所有这些都决定了数字特征在概率论与数理统计中的重要地位.

特征函数与分布函数一一对应，它虽然不像分布函数那样有直观的概率意义，但却有好得多的分析性质，因此它是解决某些分布问题的有力工具，特别在处理独立随机变量和的分布问题上占有极重要的地位. 关于特征函数与分布函数关系的另一些方面，将在下章中进一步讨论.

数字特征与特征函数都可以看作是分布函数的某种变换，而母函数则是概率论中引进的第一个变换，它在概率论发展史上有很大作用. 到了现在，母函数以它的简明性以及在某些场合的有力应用而保持了它一定的地位.

我们用特征函数作为工具研究了多元正态分布，建立了它的许多重要性质，这些良好性质从一个方面决定了正态分布在概率论与数理统计中的重要地位，正态分布重要性的另一方面——常

见性——将在下章中得到深入的讨论.

最近几十年来，信息，信息的传输、变换与处理的概念正在渗入各种各样的学科，大大加深了人类对客观世界的认识. 信息的概念与随机试验、概率分布等概率论概念密切相关，作为不肯定性度量的熵可以看作是概率空间的某种数字特征，因此我们对这些重要概念作了初步的介绍.

本章的作用是承上启下. 本章引进的两个基本概念——数字特征与特征函数——将在极限定理的研究中起重要作用.

习 题 四

1. 设随机变量 ξ 只取非负整数值，其概率为 $P\{\xi=k\}=\dfrac{a^k}{(1+a)^{k+1}}$，$a>0$ 是常数，试求 $E\xi$ 及 $D\xi$.

2. 袋中有 k 号的球 k 只，$k=1, 2, \cdots, n$，从中摸出一球，求所得号码的数学期望.

3. 随机变量 μ 取非负整数值 $n\geqslant 0$ 的概率为 $p_n=A\dfrac{B^n}{n!}$，已知 $E\mu=a$，试决定 A 与 B.

4. (1) 证明关于示性函数成立如下公式：

$$\mathbf{1}_{\overline{A}}=1-\mathbf{1}_A;\quad \mathbf{1}_{AB}=\mathbf{1}_A\cdot\mathbf{1}_B;\quad \mathbf{1}_{A\cup B}=1-\mathbf{1}_{\overline{A}\,\overline{B}}$$

(2) 利用示性函数导出概率加法公式 (1.5.6).

(3) 有五个队参加的比赛中，每个队与别的队都比赛一场，若每场比赛参赛双方各有 50%赢的机会，试求整个比赛既没有不败的队也没有不胜的队的概率.

5. 若事件 A 在第 i 次试验中出现的概率为 p_i，设 μ 是事件 A 在起初 n 次独立试验中的出现次数，试求 $E\mu$ 及 $D\mu$.

6. 一袋中含有 a 只白球，b 只黑球，从中摸出 c 只 ($c\leqslant a+b$)，求摸出白球数 μ 的数学期望.

7. 袋中有 n 张卡片，分别记上号码 1, 2, $\cdots$, n，从中有放回地抽出 k 张卡片来，求所得号码之和 μ 的数学期望及方差.

8. 某城市共有 N 辆汽车，车牌号从 1 到 N，若随机地（可重复）记下 n 辆车的车牌号，其最大号码为 ξ，求 $E\xi$.

9. 试证：若取非负整数值的随机变量 ξ 的数学期望存在，则

$$E\xi=\sum_{k=1}^{\infty}P\{\xi\geqslant k\}$$

10. 若随机变量 ξ 的分布函数为 $F(x)$，试证：

$$E\xi=\int_0^{\infty}[1-F(x)]\mathrm{d}x-\int_{-\infty}^{0}F(x)\mathrm{d}x$$

特别地，若 ξ 取非负值，则

$$E\xi=\int_0^{\infty}[1-F(x)]\mathrm{d}x$$

11. 若随机变量 ξ 服从拉普拉斯分布，其密度函数为

$$p(x)=\frac{1}{2\lambda}\mathrm{e}^{-|x-\mu|/\lambda},-\infty<x<\infty,\lambda>0$$

试求 $E\xi$ 及 $D\xi$.

12. 若分子的速度的分布密度函数由麦克斯韦分布律给出：

$$p(x)=\begin{cases}\dfrac{4x^2}{\alpha^3\sqrt{\pi}}\mathrm{e}^{-x^2/\alpha^2}, & x>0\\ 0, & x\leqslant 0\end{cases}$$

其中 $\alpha>0$ 是常数，试求分子的平均速度和平均动能（假定分子的质量等于 m).

*13. 若 ξ_1，ξ_2 相互独立，均服从 $N(a,\sigma^2)$,试证：

$$E\max(\xi_1,\xi_2)=a+\frac{\sigma}{\sqrt{\pi}}$$

14. 设 $f(x)$ $(0\leqslant x<\infty)$ 是单调非降函数，且 $f(x)>0$. 对随机变量 ξ，若 $Ef(|\xi|)<\infty$，则对任意 $x>0$，$P\{|\xi|\geqslant x\}\leqslant\dfrac{1}{f(x)}Ef(|\xi|)$.

15. 若 ξ_1，ξ_2，…，ξ_n 为正的独立随机变量，服从相同分布，密度函数为 $f(x)$，试证

$$E\left(\frac{\xi_1+\xi_2+\cdots+\xi_k}{\xi_1+\xi_2+\cdots+\xi_n}\right)=\frac{k}{n}$$

16. 袋中装有 N 只球，但其中白球数为随机变量，只知其数学期望为 n，试证从该袋中摸一球得到白球的概率为 $\dfrac{n}{N}$.

17. 甲袋中装有 a 只白球 b 只黑球，乙袋中装有 α 只白球 β 只黑球，现从甲袋中摸出 c $(c\leqslant a+b)$ 只球放入乙袋中，求从乙袋中再摸一球而为白球的概率.

*18. 袋中有 a 只白球 b 只黑球，每次摸出一球后总是放入一只白球，这样进行了 n 次之后，再从袋中摸一只球，求它是白球的概率.

*19. 甲袋中有 a 只白球 b 只黑球，乙袋中有 c 只白球 d 只黑球，从两袋中各摸出一球，并交换放入另一袋中，这样做了 n 次之后，再从甲袋中摸出一球，求这球是白球的概率.

20. 现有 n 个袋子，各装有 a 只白球 b 只黑球，先从第一个袋子中摸出一球，记下颜色后就把它放入第二个袋子中，再从第二个袋子中摸出一球，记下颜色后就把它放入第三个袋子中，照这样办法依次摸下去，最后从第 n 个袋子中摸出一球并记下颜色，若在这 n 次摸球中所摸得的白球的总数为 S_n，试求 ES_n.

21. 在物理实验中，为测量某物体的重量，通常要重复测量多次，最后再把测量记录的平均值作为该物体的重量，试说明这样做的道理.

*22. 若 ξ_1，ξ_2，…，ξ_n 是独立随机变量，$D\xi_i=\sigma_i^2$，试找"权" a_1，a_2，…，$a_n\left(\text{它们满足}\sum_{i=1}^{n}a_i=1\right)$，使 $\sum_{i=1}^{n}a_i\xi_i$ 的方差最小.

23. 设随机变量 ξ_1，ξ_2，…，ξ_{m+n} $(n>m)$ 是独立的，有相同的分布并且有有限的方差，试求 $S=\xi_1+\cdots+\xi_n$ 与 $T=\xi_{m+1}+\xi_{m+2}+\cdots+\xi_{m+n}$ 两和之间的相关系数.

24. 若 ξ 的密度函数是偶函数，且 $E\xi^2<\infty$，试证 $|\xi|$ 与 ξ 不相关，但它们不相互独立.

25. 若 $(\xi,\ \eta)$ 的密度函数为

$$p(x,y)=\begin{cases}\dfrac{1}{\pi}, & x^2+y^2\leqslant 1\\ 0, & x^2+y^2>1\end{cases}$$

试验证：ξ 与 η 不相关，但它们不独立.

26. 若 $U=aX+b$，$V=cY+d$，试证 U，V 的相关系数等于 X，Y 的相关系数. 其中 $ac>0$.

*27. 若 ξ 与 η 都是只能取两个值的随机变量，试证如果它们不相关，则独立.

28. 若 ξ_1，ξ_2，ξ_3 是三个随机变量，试讨论 (1) ξ_1，ξ_2，ξ_3 两两不相关；(2) $D(\xi_1+\xi_2+\xi_3)=D\xi_1+D\xi_2+D\xi_3$；(3) $E\xi_1\xi_2\xi_3=E\xi_1\cdot E\xi_2\cdot E\xi_3$ 之间的关系.

*29. 若 ξ，η 服从二元正态分布，$E\xi=a$，$D\xi=1$，$E\eta=b$，$D\eta=1$，证明：ξ 与 η 的相关系数 $r=\cos q\pi$，其中 $q=P\{(\xi-a)(\eta-b)<0\}$.

*30. 设 (ξ, η) 服从二元正态分布，$E\xi=E\eta=0$，$D\xi=D\eta=1$，$r_{\xi\eta}=r$，试证

$$E\max(\xi,\eta)=\sqrt{\frac{1-r}{\pi}}$$

31. 设 ξ，η 独立，均具有相同分布 $N(a, \sigma^2)$，试求 $p\xi+q\eta$ 与 $u\xi+v\eta$ 的相关系数.

32. **帕累托**（Pareto）**分布**的密度函数为

$$p(x)=\begin{cases} rA^r\dfrac{1}{x^{r+1}}, & x\geqslant A \\ 0, & x<A \end{cases}$$

这里 $r>0$，$A>0$. 试指出这分布具有 p 阶矩，当且仅当 $p<r$.

33. 若 ξ 的密度函数为

$$p(x)=\begin{cases} \dfrac{1}{2|x|(\ln|x|)^2}, & |x|>e \\ 0, & \text{其它} \end{cases}$$

试证对于任何 $\alpha>0$，$E|\xi|^{\alpha}=\infty$.

34. 若 ξ 服从 $N(a, \sigma^2)$，试求 $E|\xi-a|^k$. 其中 k 为正整数.

*35. 记 $a_k=E|\xi|^k$，若 $a_n<\infty$，试证 $\sqrt[k]{a_k}\leqslant\sqrt[k+1]{a_{k+1}}$，$k=1, 2, \cdots, n-1$.

36. 甲袋中装有 5 只白球，7 只黑球，3 只红球，乙袋中装有 4 只白球，4 只黑球，7 只红球，试问从哪一个袋中取出一只球有较大不肯定性？

37. 试求几何分布的熵.

38. 试求二项分布的熵.

39. 若以 α 及 β 分别记二进制信道的输入及输出，已知 $P\{\alpha=1\}=p$，$P\{\alpha=0\}=1-p$，$P\{\beta=1|\alpha=1\}=q$，$P\{\beta=0|\alpha=1\}=1-q$，$P\{\beta=1|\alpha=0\}=r$，$P\{\beta=0|\alpha=0\}=1-r$，试求输出中含有输入的信息量.

**40. 在12只金属球中，混有一只假球，并且不知道它是比真球轻还是重，用没有砝码的天平来称这些球.（1）试问至少需要称多少次才能查出这个假球并确定它是比真球轻或重.（2）给出一种称球方案.

41. 试用母函数法求巴斯卡分布的数学期望及方差.

42. 设 ξ 是一个母函数为 $P(s)$ 的随机变量，试求下列各概率所对应的母

函数：

（1）$P\{\xi>n\}$；　　（2）$P\{\xi=2n\}$.

**43. 在伯努利试验中，若试验次数 ν 是随机变量，试证成功的次数与失败的次数这两个随机变量独立的充要条件是 ν 服从泊松分布.

*44. 设 $\{\xi_k\}$ 是一串独立的整值随机变量序列，具有相同概率分布，考虑和 $\eta=\xi_1+\xi_2+\cdots+\xi_\nu$，其中 ν 是随机变量，它与 $\{\xi_k\}$ 相互独立，试用（1）母函数法；（2）直接计算证明

$$E\eta=E\nu\cdot E\xi_k,D\eta=E\nu\cdot D\xi_k+D\nu\cdot(E\xi_k)^2$$

45. 某公共汽车站在 $[0,t]$ 中来到的乘客批数 μ 服从参数为 λt 的泊松分布，而每批来到的乘客数是随机变量，来 n 个的概率为 p_n，$n=0,1,2,\cdots$ 试求 $[0,t]$ 中来到乘客数 η 的母函数及数学期望.

46. 试用母函数法证明二项分布及泊松分布的再生性.

*47. 若分布函数 $F(x)=1-F(-x+0)$ 成立；则称它是**对称的**；试证分布函数对称的充要条件是它的特征函数是实的偶函数.

48. 试求 $[0,1]$ 均匀分布的特征函数.

*49. 一般柯西分布的密度函数为 $p(x)=\frac{1}{\pi}\cdot\frac{\lambda}{\lambda^2+(x-\mu)^2}$，$\lambda>0$，试证它的特征函数为 $e^{i\mu t-\lambda|t|}$，利用这个结果证明柯西分布的再生性.

50. 若随机变量 ξ 服从柯西分布，$\mu=0$，$\lambda=1$，而 $\eta=\xi$，试证关于特征函数成立着

$$f_{\xi+\eta}(t)=f_\xi(t)\cdot f_\eta(t)$$

但是 ξ 与 η 并不独立.

51. 设 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立且均服从同一柯西分布，试证：$\frac{1}{n}(\xi_1+\xi_2+\cdots+\xi_n)$ 与 ξ_1 同分布.

52. 若 $\xi_1,\xi_2,\cdots,\xi_n$ 相互独立且服从相同分布 $N(0,1)$，试证 $\eta=\sum_{i=1}^{n}\xi_i^2$ 服从参数为 n 的 χ^2 分布，并说明 χ^2 分布也有再生性.

*53. 求证：对于任何实值特征函数 $f(t)$，以下两个不等式成立：

$$1-f(2t)\leqslant 4(1-f(t))$$

$$1+f(2t)\geqslant 2(f(t))^2$$

*54. 求证，如果 $f(t)$ 是相应于分布函数 $F(x)$ 的特征函数，则对于任何 x 值恒成立

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}f(t)\mathrm{e}^{-\mathrm{i}xt}\mathrm{d}t=F(x+0)-F(x-0)$$

55. 随机变量 ξ 的特征函数为 $f(t)$，且它的 n 阶矩存在，令

$$\chi_k=\frac{1}{\mathrm{i}^k}\left[\frac{\mathrm{d}^k}{\mathrm{d}t^k}\log f(t)\right]_{t=0},\quad k\leqslant n$$

称 χ_k 为随机变量 ξ 的 **k 阶半不变量**. 试证 $\eta=\xi+b$（b 是常数）的 k（$k>1$）阶半不变量等于 χ_k.

56. 试求出半不变量与原点矩之间的关系式.

57. 若（ξ，η）服从二元正态分布，试求 $\zeta=\xi+\eta$ 的分布密度函数.

58. 设 ξ_1，ξ_2，…，ξ_n 相互独立，具有相同分布 N（a，σ^2），试求 $\boldsymbol{\xi}=\begin{pmatrix}\xi_1\\ \xi_2\\ \vdots\\ \xi_n\end{pmatrix}$ 的分布，并写出它的数学期望及协方差阵. 再求 $\bar{\xi}=\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 的分布密度.

59. 若 $\boldsymbol{\xi}$ 服从二元正态分布 N（$\mathbf{0}$，$\boldsymbol{\Sigma}$），其中

$$\boldsymbol{\Sigma}=\begin{pmatrix}4 & 2\\ 2 & 1\end{pmatrix}$$

试找出矩阵 $\boldsymbol{A}$，使 $\boldsymbol{\xi}=\boldsymbol{A}\eta$，且要求 η 服从非退化的正态分布，并求 η 的密度函数.

60. 证明：在正交变换下，多元正态分布的独立、同方差性不变.

第五章　极 限 定 理

§1. 伯努利试验场合的极限定理

一、问题的提出

在第一章中我们已经指出：人们在长期实践中发现，虽然个别随机事件在某次试验中可以出现也可以不出现，但是在大量重复试验中却呈现出明显的规律性，即一个随机事件出现的频率在某个固定数的附近摆动，这就是所谓“频率稳定性”．对于这点，至今为止，我们没有给予理论上的说明．

数学上怎样来描述在一定条件下的大量重复试验呢？我们在第二章已经建立了伯努利试验这一概率模型，并指出它可以作为在一定条件下的重复试验的数学模型．在伯努利试验中，各次试验是相互独立的，在每次试验中，我们所关心的事件 A 出现的概率 $P(A)=p$ 保持不变，这些特征可以看作是从数学角度把“在一定条件下”、“重复试验”等等用语的涵义加以明确化．

在伯努利试验中，若以 μ_n 记 n 次试验中 A 出现的次数，则 $\frac{\mu_n}{n}$ 便是在这 n 次试验中事件 A 出现的频率，所谓频率稳定性无非是指当试验次数 n 增大时，频率 $\frac{\mu_n}{n}$ 接近于某个固定的常数．

这个固定的常数就是事件 A 在一次试验中发生的概率．由此可见，讨论频率 $\frac{\mu_n}{n}$ 的极限行为是理解概率论中最基本的概念——概率，所不可缺少的．正是这个缘故，在概率论的发展史上，极限定理的研究一直占重要地位，而它的发源地就是伯努利试验这个概型．

从前几章的讨论中我们知道，μ_n 是随机变量，它服从二项分布

$$P\{\mu_n=k\}=\binom{n}{k}p^k q^{n-k},\quad k=0,\ 1,\ 2,\ \cdots,\ n$$

其数学期望 $E\mu_n=np$，方差 $D\mu_n=npq$. 这在一定程度上帮助我们进一步了解了频率$\frac{\mu_n}{n}$的性质. 但是我们更需要知道的是 n 很大时 μ_n 或$\frac{\mu_n}{n}$的性质.

显然，当 n 很大时，μ_n 一般也很大，所以直接研究 μ_n 不很恰当，还是研究频率$\frac{\mu_n}{n}$为宜. 因为 $E\left(\frac{\mu_n}{n}\right)=p$，$D\left(\frac{\mu_n}{n}\right)=\frac{pq}{n}$，所以当 $n\to\infty$时，频率的数学期望保持不变，而方差则趋于 0. 我们知道方差为 0 的随机变量是常数，于是我们自然预期频率将趋于常数 p（即事件 A 发生的概率）. 但是频率$\frac{\mu_n}{n}$是随机变量，关于它的极限又将采用何种提法呢?

一种提法是：当 n 足够大时，频率$\frac{\mu_n}{n}$与概率 p 有较大偏差的概率很小. 用数学语言来讲，就是要证明：对于任意 $\varepsilon>0$,

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}=0 \tag{5.1.1}$$

或者它的等价的式子成立，即

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=1 \tag{5.1.2}$$

历史上，伯努利第一个研究了这种类型的极限定理，在 1713 年发表的论文中（这是概率论的第一篇论文!），他建立了 (5.1.1)，这是一大类概率论极限定理——大数定律 (law of large numbers）中的第一个.

关于频率接近于概率还有其他提法，譬如博雷尔建立了

$$P\left\{\lim_{n\to\infty}\frac{\mu_n}{n}=p\right\}=1 \tag{5.1.3}$$

从而开创了另一种形式的极限定理——强大数定律的研究.

本节只讨论伯努利大数定律,关于强大数定律的讨论将在§4中进行.

为了研究 μ_n 的极限行为，可以讨论它的分布 $P\{\mu_n<x\}$ 的变化情况. 但是由于 $E\mu_n=np, D\mu_n=npq$,因此对于固定的 x 来考虑 $P\{\mu_n<x\}$ 的极限不会有多大意义,因为它将趋于0,所以通常改为研究“标准化”的随机变量

$$\zeta_n=\frac{\mu_n-np}{\sqrt{npq}} \tag{5.1.4}$$

的分布函数

$$P\{\zeta_n<x\}$$

的极限行为，由 ζ_n 的分布函数不难求得 μ_n 的分布函数.

关于上述分布，已证明它的极限分布是正态分布 N（0，1），即

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t \tag{5.1.5}$$

这个结果最早由法国数学家棣莫弗（De Moivre）建立，他对 $p=\frac{1}{2}$ 证得了上述结果. 后来，拉普拉斯推广到 $0<p<1$ 的一般场合，那是另一类概率论极限定理——中心极限定理中的第一个.

下面将会看到极限定理（5.1.5）的研究直接联系到大 n 场合的二项分布的计算，这就顺便解决了第二章遗留下来的一些计算问题.

形如（5.1.5）的收敛于正态分布的极限定理的研究，在长达两个世纪的时期内成了概率论研究的中心课题，因此得到了中心极限定理的名称.

经过长期的研究，人们认识到，μ_n 具有性质（5.1.1）及（5.1.5）是由于它是独立随机变量之和，事实上，若令

$$\xi_i=\begin{cases}1，第 i 次试验出现 A\\0，第 i 次试验不出现 A\end{cases} \tag{5.1.6}$$

则

$$\mu_n=\xi_1+\xi_2+\cdots+\xi_n \tag{5.1.7}$$

这里 ξ_1，ξ_2，…，ξ_n 是相互独立的. 以后我们将会看到，对于一般的随机变量 ξ_i，$i=1$，2，…，n，也可以研究它们的和的极限定理，并且在一定条件下，这个和也具有类似于 μ_n 的性质，关于这些问题的研究就成了本章的主要内容.

为了叙述方便起见，我们引进如下定义.

定义 5.1.1 若 ξ_1，ξ_2，…，ξ_n，…是随机变量序列，令

$$\eta_n=\frac{\xi_1+\xi_2+\cdots+\xi_n}{n} \tag{5.1.8}$$

如果存在这样的一个常数序列 a_1，a_2，…，a_n，…，对任意的 $\varepsilon>0$，恒有

$$\lim_{n\to\infty}P\{|\eta_n-a_n|<\varepsilon\}=1 \tag{5.1.9}$$

则称序列 $\{\xi_n\}$ 服从**大数定律**（或**大数法则**）.

在以后的讨论中，我们几乎总是假定 ξ_1，ξ_2，…，ξ_n，…是独立随机变量序列，显然，伯努利大数定律是一般大数定律的一种特殊场合.

关于中心极限定理，我们总是对独立随机变量序列 ξ_1，ξ_2，…，ξ_n，…进行讨论，假定 $E\xi_i$ 及 $D\xi_i$ 存在，令

$$\zeta_n=\frac{\sum_{i=1}^{n}\xi_i-\sum_{i=1}^{n}E\xi_i}{\sqrt{\sum_{i=1}^{n}D\xi_i}} \tag{5.1.10}$$

我们的目的是寻找使

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t \tag{5.1.11}$$

成立的条件.

一般，若独立随机变量序列 ξ_1，ξ_2，…，ξ_n，…的标准化和 ζ_n 使（5.1.11）成立，则我们称 $\{\xi_i\}$ 服从**中心极限定理**（central limit theorem）.

显然，棣莫弗—拉普拉斯极限定理是中心极限定理的特例.

二、伯努利大数定律

我们将证明一个比伯努利大数定律更强的命题.

1. 车贝晓夫大数定律

车贝晓夫大数定律 设 ξ_1，ξ_2，…，ξ_n，…是由两两不相关的随机变量所构成的序列，每一随机变量都有有限的方差，并且它们有公共上界

$$D\xi_1 \leqslant C,\ D\xi_2 \leqslant C,\ \cdots,\ D\xi_n \leqslant C,\ \cdots$$

则对任意的 $\varepsilon > 0$，皆有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}E\xi_k\right| < \varepsilon\right\} = 1 \qquad (5.1.12)$$

［证明］ 因为 $\{\xi_k\}$ 两两不相关，故

$$D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right) = \frac{1}{n^2}\sum_{k=1}^{n}D\xi_k \leqslant \frac{C}{n}$$

再由车贝晓夫不等式得到

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}E\xi_k\right| < \varepsilon\right\} \geqslant 1 - \frac{D\left(\frac{1}{n}\sum_{k=1}^{n}\xi_k\right)}{\varepsilon^2} \geqslant 1 - \frac{C}{n\varepsilon^2}$$

所以

$$1 \geqslant P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}\xi_k - \frac{1}{n}\sum_{k=1}^{n}E\xi_k\right| < \varepsilon\right\} \geqslant 1 - \frac{C}{n\varepsilon^2}$$

于是，当 $n\to\infty$ 时有（5.1.12），因此定理得证.

这个结果在 1866 年被俄国数学家车贝晓夫所证明，它是关于大数定律的一个相当普遍的结论，许多大数定律的古典结果是它

的特例；此外，证明这个定律所用的方法后来称为矩法，也很有创造性，在这基础上发展起来的一系列不等式是研究各种极限定理的有力工具.

马尔可夫（Марков）注意到在车贝晓夫的论证中，只要

$$\frac{1}{n^2}D\Big(\sum_{k=1}^{n}\xi_k\Big)\to 0 \tag{5.1.13}$$

则大数定律就能成立，通常称条件（5.1.13）为**马尔可夫条件**.

马尔可夫大数定律 对于随机变量序列 $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ 若（5.1.13）成立，则对任意 $\varepsilon>0$，均有（5.1.12）.

车贝晓夫大数定律显然可由马尔可夫大数定律推出；更重要的是马尔可夫大数定律已经没有任何关于独立性的假定. 研究相依随机变量序列的大数定律是近代概率论的课题之一，但是这已超出我们讨论的范围.

2. 伯努利大数定律与泊松大数定律

伯努利大数定律 设 μ_n 是 n 次伯努利试验中事件 A 出现的次数，而 p 是事件 A 在每次试验中出现的概率，则对任意 $\varepsilon>0$，都有

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=1 \tag{5.1.14}$$

[证明] 定义随机变量 ξ_i 如（5.1.6），则

$$E\xi_i=p,\quad D\xi_i=pq\leqslant\frac{1}{4}$$

而

$$\frac{1}{n}\sum_{k=1}^{n}\xi_k-\frac{1}{n}\sum_{k=1}^{n}E\xi_k=\frac{\mu_n}{n}-p$$

故由车贝晓夫大数定律立刻推出伯努利大数定律. 显然，伯努利大数定律也可以通过车贝晓夫不等式直接加以证明：

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}D\left\{\frac{\mu_n}{n}\right\}=\frac{1}{n\varepsilon^2}D\xi_i\leqslant\frac{1}{4n\varepsilon^2} \tag{5.1.15}$$

历史上，伯努利大数定律是通过直接展开和繁复的计算才证得的.

泊松大数定律 如果在一个独立试验序列中，事件 A 在第 k 次试验中出现的概率等于 p_k，以 μ_n 记在前 n 次试验中事件 A 出现的次数，则对任意 $\varepsilon>0$，都有

$$\lim_{n\to\infty}P\left\{\left|\frac{\mu_n}{n}-\frac{p_1+p_2+\cdots+p_n}{n}\right|<\varepsilon\right\}=1 \qquad (5.1.16)$$

[证明] 定义 ξ_k 为第 k 次试验中事件 A 出现的次数，则

$$E\xi_k=p_k, \qquad D\xi_k=p_k\ (1-p_k)\leqslant\frac{1}{4}$$

再用车贝晓夫大数定律立刻可以推出（5.1.16）

泊松提出了不同于伯努利试验的另一种独立试验模型，并导出泊松分布以及证明了上述大数定律，从而奠定了他在概率论发展史上的重要地位.

3. 大数定律的重要意义

人们积累的大量经验告诉我们，具有很接近于 1 的概率的随机事件在一次试验中几乎一定要发生；同样，概率很小的事件在一次试验中可以看作是实际不可能事件. 至于概率小到何种程度才能看作是实际不可能事件则要视事件的重要性而定，有百分之一可能性含有染菌的药物是应该废弃的，但含有百分之一次品的钮扣则问题还不太大.

因此在实际工作及一般理论问题中，概率接近于 1 或 0 的事件具有重大意义，概率论的基本问题之一就是要建立概率接近于 1 或 0 的规律；特别是大量独立或弱相关因素累积结果所发生的规律. 大数定律就是这种概率论命题中最重要的一个.

观察个别现象时是连同一切个别的特性来观察的. 这些个别的特性往往蒙蔽了事物的规律性. 在大量观察中个别因素的影响将相互抵消而使总体稳定. 例如，虽然每个气体分子的运动带有很大的随机性，但是作为气体平均特征的压力、温度等却是稳定

的，大数定律说明了这种稳定性．古人用“定律”来称呼这类命题，可见其重视的程度．

伯努利大数定律建立了在大量重复独立试验中事件出现频率的稳定性，正因为这种稳定性，概率的概念才有客观意义．伯努利大数定律还提供了通过试验来确定事件概率的方法，既然频率 $\frac{\mu_n}{n}$ 与概率 p 有较大偏差的可能性很小，那么我们便可以通过做试验确定某事件发生的频率并把它作为相应概率的估计．这类方法称为**参数估计**，它是数理统计中的主要研究课题之一，参数估计的重要理论基础之一就是大数定律．

三、棣莫弗—拉普拉斯极限定理

大数定律只断言 $P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}$ 当 $n\to\infty$ 时趋于 0，也即 $\frac{\mu_n}{n}$ 接近于 p，而棣莫弗—拉普拉斯极限定理则给出 μ_n 的渐近分布的更精确表述．下面定理给出了两个结果．第一个结果提供了 $P\{\mu_n=k\}$ 的渐近表达式，这类结果一般称为**局部极限定理**．第二个结果给出了标准化随机变量 $\frac{\mu_n-np}{\sqrt{npq}}$ 的渐近分布，称为**积分极限定理**，它是一般中心极限定理的特例．这两个结果既有区别也有联系．

定理 5.1.1（棣莫弗—拉普拉斯） 若 μ_n 是 n 次伯努利试验中事件 A 出现的次数，$0<p<1$，则对任意有限区间 $[a,b]$：

(i) 当 $a\leqslant x_k\equiv\frac{k-np}{\sqrt{npq}}\leqslant b$ 及 $n\to\infty$ 时，一致地有

$$P\{\mu_n=k\}\div\left(\frac{1}{\sqrt{npq}}\cdot\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x_k^2}\right)\to 1 \qquad (5.1.17)$$

(ii) 当 $n\to\infty$ 时，一致地有

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\to\int_a^b\varphi(x)\mathrm{d}x \qquad (5.1.18)$$

其中 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}(-\infty<x<\infty)$．

［证明］　先证局部极限定理，我们将给出一个比（5.1.17）更为精确的渐近式.

因 x_k 只能在有限区间 $[a, b]$ 中取值，故当 $n\to\infty$时，

$$k=np+x_k\sqrt{npq}\to\infty \tag{5.1.19}$$

$$j\equiv n-k=nq-x_k\sqrt{npq}\to\infty \tag{5.1.20}$$

由斯特林（Stirling）公式：

$$m!=\sqrt{2\pi m}m^m\mathrm{e}^{-m}\mathrm{e}^{\theta_m}\quad\left(0<\theta_m<\frac{1}{12m}\right) \tag{5.1.21}$$

可知

$$\begin{aligned}P\{\mu_n=k\}&=\frac{n!}{k!j!}p^kq^j\\&=\frac{\sqrt{2\pi n}n^n\mathrm{e}^{-n}}{\sqrt{2\pi k}k^k\mathrm{e}^{-k}\sqrt{2\pi j}j^j\mathrm{e}^{-j}}p^kq^j\mathrm{e}^{\theta_n-\theta_k-\theta_j}\\&=\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{npq}}\left(\frac{np}{k}\right)^{k+\frac{1}{2}}\left(\frac{nq}{n-k}\right)^{n-k+\frac{1}{2}}\mathrm{e}^{\theta}\end{aligned} \tag{5.1.22}$$

其中 $\theta=\theta_n-\theta_k-\theta_{n-k}$，因此

$$|\theta|<\frac{1}{12}\left(\frac{1}{n}+\frac{1}{k}+\frac{1}{n-k}\right) \tag{5.1.23}$$

由（5.1.19）及（5.1.20）知

$$\frac{k}{np}=1+x_k\sqrt{\frac{q}{np}},\quad\frac{n-k}{nq}=1-x_k\sqrt{\frac{p}{nq}} \tag{5.1.24}$$

我们将利用展开式

$$\ln(1+x)=x-\frac{x^2}{2}+\frac{x^3}{3}-\frac{x^4}{4}+\cdots \tag{5.1.25}$$

来对它们进行估计. 这个展开式当$-1<x\leqslant1$时收敛，但只有对绝对值很小的那种 x 值才收敛得快. 现在，当 n 充分大时，$x_k\sqrt{\frac{q}{np}}$ 及 $x_k\sqrt{\frac{p}{nq}}$都很小（因此，当 $p=0$ 及 $p=1$ 时不能用；此外，当 p 或q 很小时，渐近展开式引起的误差也较大，这时我们已推荐用泊

松逼近公式）．所以

$$\ln\left(\sqrt{2\pi npq}P\{\mu_n=k\}\right)=\theta-\left(k+\frac{1}{2}\right)\ln\frac{k}{np}-\left(n-k+\frac{1}{2}\right)\ln\frac{n-k}{nq}$$

$$=\theta-\left(np+x_k\sqrt{npq}+\frac{1}{2}\right)\ln\left(1+x_k\sqrt{\frac{q}{np}}\right)$$

$$-\left(nq-x_k\sqrt{npq}+\frac{1}{2}\right)\ln\left(1-x_k\sqrt{\frac{p}{nq}}\right)$$

$$=\theta-\left(np+x_k\sqrt{npq}+\frac{1}{2}\right)$$

$$\cdot\left[x_k\sqrt{\frac{q}{np}}-\frac{x_k^2q}{2np}+\frac{x_k^3q\sqrt{npq}}{3n^2p^2}-\frac{x_k^4q^2}{4n^2p^2}+\cdots\right]$$

$$-\left(nq-x_k\sqrt{npq}+\frac{1}{2}\right)$$

$$\cdot\left[-x_k\sqrt{\frac{p}{nq}}-\frac{x_k^2p}{2nq}-\frac{x_k^3p\sqrt{npq}}{3n^2q^2}-\frac{x_k^4p^2}{4n^2q^2}+\cdots\right]$$

$$=\theta-\frac{x_k^2}{2}+\frac{q-p}{6\sqrt{npq}}(x_k^3-3x_k)$$

$$+\frac{1}{12npq}\left[3(p^2+q^2)x_k^2-(p^3+q^3)x_k^4\right]+o\left(\frac{1}{n}\right)\tag{5.1.26}$$

因此

$$P\{\mu_n=k\}=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{npq}}\exp\left\{-\frac{x_k^2}{2}+\frac{(q-p)(x_k^3-3x_k)}{6\sqrt{npq}}+o\left(\frac{1}{n}\right)\right\}$$

$$=\frac{1}{\sqrt{2\pi}}\cdot\frac{1}{\sqrt{npq}}e^{-\frac{x_k^2}{2}}\left[1+\frac{(q-p)(x_k^3-3x_k)}{6\sqrt{npq}}+o\left(\frac{1}{n}\right)\right]\tag{5.1.27}$$

取其第一项即得(5.1.17)，因此我们已证得了局部极限定理．显然，我们得到了更精确的估计式。又知当 $p=q$ 及 $x_k^3-3x_k=0$ 时，近似效果尤佳．

下面转入证明积分极限定理.

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}$$
$$=P\{np+a\sqrt{npq}\leqslant\mu_n<np+b\sqrt{npq}\}$$
$$=\sum_{k=k_1}^{k_2}P\{\mu_n=k\} \tag{5.1.28}$$

上式中 k_1 为不小于 $np+a\sqrt{npq}$ 的最小整数，k_2 为小于 $np+b\sqrt{npq}$ 的最大整数，由局部极限定理知当 n 充分大时，对任给 $\varepsilon>0$，有

$$P\{\mu_n=k\}=\frac{1}{\sqrt{npq}}(\varphi(x_k)+\varepsilon_k),\ |\varepsilon_k|<\varepsilon$$
$$(k=k_1,\ k_1+1,\ \cdots,\ k_2)$$

代入（5.1.28）式得到

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}$$
$$=\sum_{k=k_1}^{k_2}\frac{1}{\sqrt{npq}}\varphi(x_k)+\sum_{k=k_1}^{k_2}\frac{\varepsilon_k}{\sqrt{npq}}$$

因为有

$$\left|\sum_{k=k_1}^{k_2}\frac{\varepsilon_k}{\sqrt{npq}}\right|\leqslant\frac{1}{\sqrt{npq}}(k_2-k_1+1)\varepsilon\leqslant\frac{(b-a)\sqrt{npq}+1}{\sqrt{npq}}\varepsilon$$

故当 $n\to\infty$ 时，注意 x_k 的增量为 $\frac{1}{\sqrt{npq}}$，就得到

$$P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\to\int_a^b\varphi(x)\mathrm{d}x$$

定理到此完全证毕.

利用 $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mathrm{e}^{-x^2/2}\mathrm{d}x=1$，不难证明在积分极限定理中，当 $a=-\infty$，$b=+\infty$ 时仍然成立.

四、棣莫弗—拉普拉斯极限定理的一些应用

棣莫弗—拉普拉斯极限定理虽然是作为二项分布的近似而推导出来的，但是它的重要性远远超出数值计算的范围．对它的各种推广形式的深入讨论将在以后几节进行，这里我们先介绍一些它的具体应用，其中有些解决了第一、二章遗留的问题．

［导出伯努利大数定律］ 积分极限定理断言，μ_n 的分布渐近于正态分布 $N(np, npq)$，因此伯努利试验中事件 A 出现的频率 $\frac{\mu_n}{n}$ 的分布渐近于正态分布 $N\left(p, \frac{pq}{n}\right)$．这里可以想象，当 $n\to\infty$ 时，$\frac{\mu_n}{n}$ 的分布会收敛于退化分布

$$I(x-p)=\begin{cases}0, x\leqslant p\\ 1, x>p\end{cases}$$

这正是伯努利大数定律所确定的事实．下面来严格证明．

因为 ε 固定，对任意正数 l，只要 n 足够大，就有

$$l\sqrt{npq}<\varepsilon n$$

因此

$$\left\{\left|\frac{\mu_n-np}{\sqrt{npq}}\right|<l\right\}\subset\left\{\left|\frac{\mu_n-np}{n}\right|<\varepsilon\right\}$$

所以对大的 n，

$$P\left\{\left|\frac{\mu_n-np}{\sqrt{npq}}\right|<l\right\}\leqslant P\left\{\left|\frac{\mu_n-np}{n}\right|<\varepsilon\right\}$$

由积分极限定理，当 $n\to\infty$ 时，上式左边收敛于

$$\frac{1}{\sqrt{2\pi}}\int_{-l}^{l}\mathrm{e}^{-x^2/2}\mathrm{d}x$$

对于任给 $\delta>0$，可以选 l，使得上面积分值大于 $1-\delta$，因此对充分大的 n 有

$$P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}>1-\delta$$

这就证明了伯努利大数定律，从这里也可看出，积分极限定理比大数定律更精确.

[用频率估计概率时误差的估计] 由积分极限定理

$$P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}=P\left\{-\varepsilon\sqrt{\frac{n}{pq}}<\frac{\mu_n-np}{\sqrt{npq}}<\varepsilon\sqrt{\frac{n}{pq}}\right\}$$

$$\approx\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-\Phi\left(-\varepsilon\sqrt{\frac{n}{pq}}\right)=2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-1 \tag{5.1.29}$$

这个关系式可用来解决许多计算问题.

第一类问题是已知 n，p，ε，求概率 $P\left\{\left|\frac{\mu_n}{n}-p\right|<\varepsilon\right\}$；这时只要利用 (5.1.29) 并查正态分布函数 $\Phi(x)$ 的数值表就可解决，这类问题在二项分布计算中经常会遇到.

第二类问题是要使$\frac{\mu_n}{n}$与 p 的差异不大于定数 ε 的概率不小于预先给的数 β，问最少应做多少次试验？这时只需要求满足下式的最小 n，

$$2\Phi\left(\varepsilon\sqrt{\frac{n}{pq}}\right)-1\geqslant\beta$$

这也可通过查表求得.

第三类问题已知 n 及 β，求 ε，这类问题是在进行误差估计时提出来的. 解法如下：先找 x_β 使

$$2\Phi(x_\beta)-1=\beta$$

这时

$$\varepsilon=x_\beta\sqrt{\frac{pq}{n}}$$

即为所求. 若 p 不知道，则利用 $pq\leqslant\frac{1}{4}$，有下列估计式

$$\varepsilon\leqslant\frac{x_\beta}{2\sqrt{n}}$$

这类估计在蒙特卡洛方法中很有用处.

[局部极限定理在二项分布计算中的应用] 局部极限定理给出

$$\binom{n}{k} p^k q^{n-k} \div \left(\frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{k-np}{\sqrt{npq}}\right)^2} \right) \to 1 \quad (5.1.30)$$

这个事实由图 27 表示出来，其中阶梯函数给出概率 $\binom{n}{k} p^k q^{n-k}$，而粗线则给出对应的正态分布密度函数曲线.

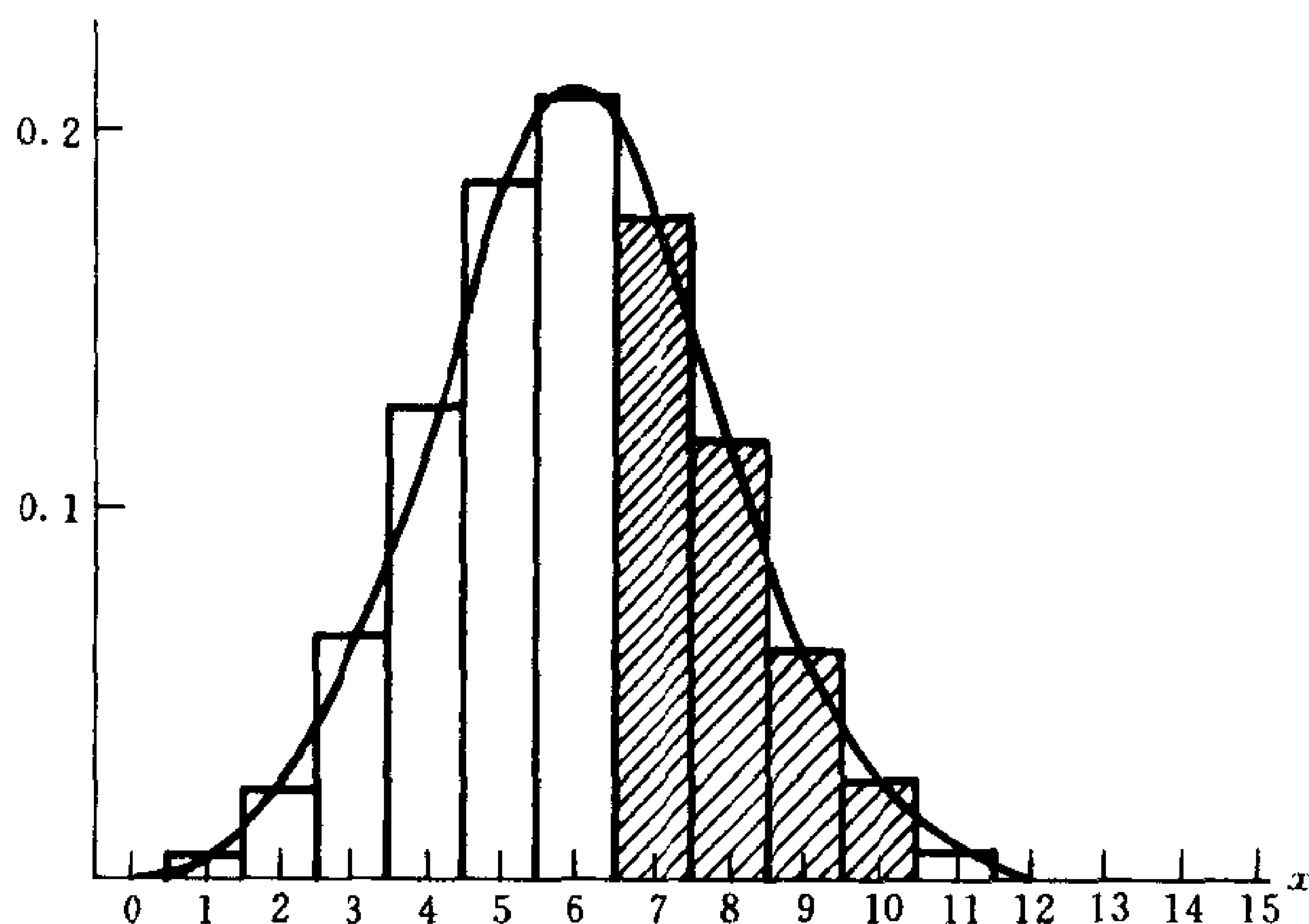


图 27 二项分布的正态逼近

我们来利用局部极限定理对 $\binom{n}{k} p^k q^{n-k}$ 进行近似计算.

因为 $$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad -\infty < x < \infty \quad (5.1.31)$$

有专门数值表，当 n 较大时，对二项分布的计算可用下列近似式：

$$\binom{n}{k} p^k q^{n-k} \approx \frac{1}{\sqrt{npq}} \varphi(x_k) \quad (5.1.32)$$

其中 $x_k = \frac{k-np}{\sqrt{npq}}$. 用这种方法计算当然会有误差，但实践证明（也能利用 (5.1.27) 作理论分析），当 p 不太接近于 0 或 1，而 n 又不太小时，都能得到良好的结果.

下面就用这种办法来解决第二章§4中提出的一些二项分布计算问题.

例3中要求计算 $b(5;500,0.01)=\binom{500}{5}(0.01)^5\cdot(0.99)^{495}$，利用(5.1.32)

$$
\begin{aligned}
&b(5;500,0.01)\\
&\approx\frac{1}{\sqrt{500\times0.01\times0.99}}\varphi\left(\frac{5-500\times0.01}{\sqrt{500\times0.01\times0.99}}\right)\\
&=\frac{1}{\sqrt{4.95}}\varphi(0)=\frac{1}{\sqrt{4.95}}\times0.398942=0.17931
\end{aligned}
$$

第二章中已指出 $b(5;500,0.01)$ 的精确值是0.17635，这里 $p=0.01$ 很接近于0，我们用近似计算得到的数值的有效性还达到小数点后两位，由此可见用局部极限定理计算的近似程度是很好的.

再来计算例4（1），那里要求计算 $b(40;10000,0.005)$，利用(5.1.32)

$$
\begin{aligned}
&b(40;10000,0.005)\\
&\approx\frac{1}{\sqrt{10000\times0.005\times0.995}}\varphi\left(\frac{40-10000\times0.005}{\sqrt{10000\times0.005\times0.995}}\right)\\
&=\frac{1}{7.05}\varphi(1.4)=\frac{1}{7.05}\times0.1497=0.0212
\end{aligned}
$$

而精确值为0.0214，准确度甚高.

[积分极限定理在二项分布计算中的应用] 积分极限定理给出

$$
P\left\{a\leqslant\frac{\mu_n-np}{\sqrt{npq}}<b\right\}\to\Phi(b)-\Phi(a) \qquad (5.1.33)
$$

其中 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ 是正态分布函数,利用它可以计算形如 $P\{k_1\leqslant\mu_n\leqslant k_2\}$ 的概率.

因为当 n 充分大时，$b(k;n,p)$ 即使对于最可能成功次数，其数值也很小，这时讨论成功次数等于某数 k 的概率，即求 $P\{\mu_n$

$=k\}$ 便没有太大意思；相反，计算 μ_n 的值落在某一区间 $[k_1, k_2]$ 的概率倒十分重要，因此积分极限定理在二项分布近似计算中更为有用.

我们有

$$P\{k_1 \leqslant \mu_n \leqslant k_2\} = P\left\{\frac{k_1 - np}{\sqrt{npq}} \leqslant \frac{\mu_n - np}{\sqrt{npq}} \leqslant \frac{k_2 - np}{\sqrt{npq}}\right\}$$

$$\approx \Phi\left(\frac{k_2 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np}{\sqrt{npq}}\right) \tag{5.1.34}$$

细察图 27 可以看出，在用连续型分布正态分布逼近离散型分布二项分布时存在着偏差. 对应于离散值 7，8，9，10 的矩形面积，相应的该是 $6.5 < x < 10.5$ 上的曲边梯形面积，因此下面的修正公式，常能得到更好的近似效果：

$$P\{k_1 \leqslant \mu_n \leqslant k_2\} \approx \Phi\left(\frac{k_2 - np + 0.5}{\sqrt{npq}}\right) - \Phi\left(\frac{k_1 - np - 0.5}{\sqrt{npq}}\right) \tag{5.1.35}$$

当 p 不太接近 0 或 1，而 n 又不太小时，用这个近似式能得到良好的结果.

例如，我们来完成第二章 § 4 中的一些计算. 例 4（2）要计算 $P\{\mu \leqslant 70\}$，利用(5.1.35)

$$P\{\mu \leqslant 70\}$$

$$\approx \Phi\left(\frac{70 - 10000 \times 0.005 + 0.5}{\sqrt{10000 \times 0.005 \times 0.995}}\right) - \Phi\left(\frac{-10000 \times 0.005 - 0.5}{\sqrt{10000 \times 0.005 \times 0.995}}\right)$$

$$= \Phi(2.91) - \Phi(-7.16) = \Phi(2.91) - [1 - \Phi(7.16)] = 0.998$$

这里 $\Phi(7.16)$ 取为 1，此外计算中还用到 $\Phi(-x) = 1 - \Phi(x)$，这由 $\varphi(x)$ 的对称性即得.

再来计算例 5，其中 $p = \frac{1}{2}$，$n = 5.4 \times 10^{22}$.

$$P\{|\mu - np| > 2.7 \times 10^{12}\} = P\left\{\frac{|\mu - np|}{\sqrt{npq}} > \frac{2.7 \times 10^{12}}{\sqrt{5.4 \times 10^{22} \times \frac{1}{4}}}\right\}$$

$$=P\left\{\frac{|\mu-np|}{\sqrt{npq}}>2.32\times 10\right\}\approx\frac{2}{\sqrt{2\pi}}\int_{23.2}^{\infty}e^{-t^2/2}dt$$

这个数值非常小，从 $\Phi(x)$ 的数值表中不能找到，但我们可以用下面方法对它进行估计.

由于对 $z>0$ 有

$$\int_z^{\infty}e^{-t^2/2}dt<\frac{1}{z}\int_z^{\infty}te^{-t^2/2}dt=\frac{1}{z}e^{-z^2/2} \tag{5.1.36}$$

利用它可以得到

$$\frac{2}{\sqrt{2\pi}}\int_{23.2}^{\infty}e^{-t^2/2}dt<\frac{2}{\sqrt{2\pi}}\cdot\frac{1}{23.2}e^{-269.12}<10^{-100}$$

这概率非常小.

现在我们能够解决第一章§1中提出的用电问题了. 该问题可以重新叙述如下：

某车间有200台车床，它们独立地工作着，开工率各为0.6，开工时耗电各为1千瓦，问供电所至少要供给这个车间多少电力才能以99.9%的概率保证这个车间不会因供电不足而影响生产?

这里的叙述已把问题数学模型化，提出了各车床工作的独立性假定，并把所谓能正常生产的要求加以明确化，这是在用数学解决生产实际问题时必须经过的一步.

现在，这成为试验次数 $n=200$ 的伯努利试验，若把某台车床在工作看作成功，则出现成功的概率为0.6. 记某时在工作着的车床数为 ξ，则 ξ 是随机变量，服从 $p=0.6$ 的二项分布. 问题是要求 r，使

$$P\{\xi\leqslant r\}=\sum_{k=0}^{r}\binom{200}{k}(0.6)^k(0.4)^{200-k}\geqslant 0.999 \tag{5.1.37}$$

我们可以利用积分极限定理计算这个概率.

$$\sum_{k=0}^{r}\binom{200}{k}(0.6)^k(0.4)^{200-k}$$

$$\approx\Phi\left(\frac{r-200\times 0.6+0.5}{\sqrt{200\times 0.6\times 0.4}}\right)-\Phi\left(\frac{-200\times 0.6-0.5}{\sqrt{200\times 0.6\times 0.4}}\right)$$

$$=\Phi\left(\frac{r-119.5}{\sqrt{48}}\right)-\Phi\ (-17.39)\approx\Phi\left(\frac{r-119.5}{\sqrt{48}}\right)\geqslant 0.999$$

查表得

$$\frac{r-119.5}{\sqrt{48}}=3.1$$

所以

$$r=141$$

这个结果表明 $P\{\xi\leqslant 141\}\geqslant 0.999$，所以我们若供电 141 千瓦，那么由于供电不足而影响生产的可能性小于 0.001，相当于在 8 小时工作中有半分钟受影响，这在一般工厂中是允许的. 当然不同的生产单位，可能提出不同的要求，那么我们可以改变 (5.1.37) 右端的概率值，但是方法还是同样的.

现在也能对高尔顿板这个试验作理论解释了. 读者可能早已想到，近似小球高度的曲线是正态分布密度函数曲线. 这完全正确. 事实上，高尔顿板可以看作是伯努利试验的一个实验模型. 如果我们把小球碰到钉子看作是一次试验，而把从右边落下算是成功，当然从左边落下就算是失败，这时就有了一次 $p=\frac{1}{2}$ 的伯努利试验. 小球从顶端到底层共需要经过 n 排钉子，这就相当于一个 n 次伯努利试验，剩下的只是要说明为什么高度曲线会是正态分布密度函数曲线，这个问题留给读者思考.

§2. 收 敛 性

从下节开始，我们将把在伯努利试验场合建立的极限定理推广到更为一般的场合，本节为此准备了必要的概念与工具.

从上节的讨论中我们已经看到，概率论的极限定理研究的是随机变量序列与分布函数序列的某种收敛性，下面我们将给这些收敛性以明确定义并讨论它们的有关性质. 这些结果对于深入研究概率论也有着独自的重要性.

特征函数是研究极限定理的有力工具，从上章的讨论中我们

已经知道，它与分布函数互相唯一确定，本节中我们将证明这种对应还具有某种连续性，这些性质决定了特征函数在极限定理研究中的特殊地位. 顺便还得到特征函数的充要条件，这个结果在平稳随机过程的研究中有基本的重要性. 本节定理的证明都较长，初学时可以略去，后面用到的是结论.

一、分布函数弱收敛

中心极限定理讨论的是分布函数列收敛于正态分布. 事实上，在棣莫弗—拉普拉斯积分极限定理中，若记

$$F_n(x)=P\left\{\frac{\mu_n-np}{\sqrt{npq}}<x\right\}$$

则定理的结论可以表述为

$$F_n(x)\to\Phi(x)$$

这正是一个分布函数列 $\{F_n(x)\}$ 收敛于某一个分布函数 $\Phi(x)$，这种收敛对每一点 x 都成立.

这个考察对于我们引进一般分布函数列的收敛性定义很有帮助，在给出定义前，我们再来看一个例子.

［例 1］ 令

$$F_n(x)=\begin{cases}0, & x\leqslant-\dfrac{1}{n}\\[2mm] 1, & x>-\dfrac{1}{n}\end{cases} \tag{5.2.1}$$

这是一个退化分布，它可以解释为一个单位质量全部集中在 $x=-\frac{1}{n}$ 这一点的分布. 当 $n\to\infty$ 时，我们自然认为 $\{F_n(x)\}$ 应该收敛于一个单位质量全部集中在 $x=0$ 这一点的分布，即

$$F(x)=\begin{cases}0, & x\leqslant 0\\ 1, & x>0\end{cases}$$

但是，$F_n(0)\equiv1$，而 $F(0)=0$，显然，$F_n(0)\not\to F(0)$. 因此看来要求分布函数列在所有的点都收敛到极限分布函数是太严了. 上例中

不收敛的点是极限分布函数 $F(x)$的不连续点.

定义 5.2.1 对于分布函数列$\{F_n(x)\}$,如果存在一个非降函数 $F(x)$使

$$\lim_{n\to\infty} F_n(x)=F(x) \tag{5.2.2}$$

在 $F(x)$的每一连续点上都成立,则称 $F_n(x)$**弱收敛于** $F(x)$,并记为 $F_n(x)\xrightarrow{W} F(x)$.

这样得到的极限函数是一个有界的非降的函数，我们也可以选得它是左连续的，但是下例说明，它不一定是一个分布函数.

［例 2］ 取

$$F_n(x)=\begin{cases}0, & x\leqslant n\\ 1, & x>n\end{cases} \tag{5.2.3}$$

显然 $\lim\limits_{n\to\infty} F_n(x)=0$ 对一切 x 成立,但 $F(x)\equiv 0$ 不是分布函数.

当然,若已知分布函数列$\{F_n(x)\}$弱收敛于分布函数 $F(x)$及 $G(x)$,则 $F(x)=G(x)$对一切 x 成立.

我们希望能得到一个分布函数列弱收敛到一个分布函数的充要条件，为此先建立一些重要的分析结果，这些结果对一般的有界非降函数列都成立，它们的弱收敛概念类似地定义.

引理 5.2.1 设 $\{F_n(x)\}$ 是实变量 x 的非降函数列，D 是 $\mathbf{R}^1$ 上的稠密集. 若对于 D 中的所有点，序列 $\{F_n(x)\}$ 收敛于 $F(x)$，则对 $F(x)$的一切连续点 x 有

$$\lim_{n\to\infty} F_n(x)=F(x) \tag{5.2.4}$$

［证明］ 设 x 是任意点,选 $x'\in D, x''\in D$,使 $x'\leqslant x\leqslant x''$,由非降性知

$$F_n(x')\leqslant F_n(x)\leqslant F_n(x'')$$

因此

$$F(x')\leqslant \varliminf_{n\to\infty} F_n(x)\leqslant \varlimsup_{n\to\infty} F_n(x)\leqslant F(x'')$$

因为 D 在 $\mathbf{R}^1$ 上稠密,故

$$F(x-0)\leqslant\varliminf_{n\to\infty}F_n(x)\leqslant\varlimsup_{n\to\infty}F_n(x)\leqslant F(x+0)$$

所以对于 $F(x)$的连续点 x,成立(5.2.4).

下面证明海莱(Helly)的两个重要定理.

定理 5.2.1（海莱第一定理） 任一一致有界的非降函数列$\{F_n(x)\}$中必有一子序列$\{F_{n_k}(x)\}$弱收敛于某一有界的非降函数$F(x)$.

［证明］ 任取$\mathbf{R}^1$上的一个到处稠密的可数点集D,下面我们就取有理数全体，并排列为$r_1,r_2,\cdots,r_m,\cdots$对于序列$\{F_n(r_1)\}$，这是一个有界的实数序列，故必包含一收敛于某极限$G(r_1)$的子序列$\{F_{1,n}(r_1)\}$，即

$$\lim_{n\to\infty}F_{1,n}(r_1)=G(r_1)$$

现在考虑序列$\{F_{1,n}(r_2)\}$，同样由于有界性，在其中存在子序列$\{F_{2,n}(r_2)\}$收敛于某一值$G(r_2)$. 这时,同时成立着

$$\lim_{n\to\infty}F_{2,n}(r_1)=G(r_1),\quad \lim_{n\to\infty}F_{2,n}(r_2)=G(r_2)$$

继续这样做,可得序列$\{F_{m,n}(x)\}$,使

$$\lim_{n\to\infty}F_{m,n}(r_k)=G(r_k),\quad k=1,2,\cdots,m \tag{5.2.5}$$

同时成立.

这样,我们得到了$\{F_n(x)\}$的如下子序列

$$\begin{matrix}F_{1,1}(x), & F_{1,2}(x), & F_{1,3}(x), & \cdots, & F_{1,n}(x), & \cdots\\ F_{2,1}(x), & F_{2,2}(x), & F_{2,3}(x), & \cdots, & F_{2,n}(x), & \cdots\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots & \\ F_{m,1}(x), & F_{m,2}(x), & F_{m,3}(x), & \cdots, & F_{m,n}(x), & \cdots\\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots & \cdots\cdots & \end{matrix} \tag{5.2.6}$$

这里每行都是前一行的子序列，而且它们具有性质（5.2.5）. 选取这个阵列的对角线元素$F_{n,n}(x)$构成新序列$\{F_{n,n}(x)\}$，由于它是从$\{F_{1,n}(x)\}$分出来的,故$\lim\limits_{n\to\infty}F_{n,n}(r_1)=G(r_1)$. 其次,除第一项外,它是由$\{F_{2,n}(x)\}$分出来的,故$\lim\limits_{n\to\infty}F_{n,n}(r_2)=G(r_2)$. 一般地,

对任何固定的 k，皆有 $\lim\limits_{n\to\infty}F_{n,n}(r_k)=G(r_k)$，因此对一切有理数 r，

$$\lim_{n\to\infty}F_{n,n}(r)=G(r) \tag{5.2.7}$$

这里的 $G(r)$ 是定义在有理数上的函数，它也是有界与非降的.

对一切 $x\in\mathbf{R}^1$，定义

$$F(x)=\sup_{r_k\leqslant x}G(r_k)$$

这函数在一切有理数上与 $G(x)$ 相等，它显然也是有界与非降的.

由引理 5.2.1 知

$$\lim_{n\to\infty}F_{n,n}(x)=F(x) \tag{5.2.8}$$

对 $F(x)$ 的一切连续点成立，这就证明了定理 5.2.1. 通常形象地称这个定理的证明方法为**对角线法**.

极限函数 $F(x)$ 不一定左连续，但总可以改变它不连续点上的值使之左连续，这样的改变显然不影响（5.2.8）的成立.

定理 5.2.2 （海莱第二定理） 设 $f(x)$ 是 $[a,b]$ 上的连续函数，又 $\{F_n(x)\}$ 是在 $[a, b]$ 上弱收敛于函数 $F(x)$ 的一致有界非降函数序列，且 a 和 b 是 $F(x)$ 的连续点，则

$$\lim_{n\to\infty}\int_a^b f(x)\mathrm{d}F_n(x)=\int_a^b f(x)\mathrm{d}F(x)$$

［证明］ 由函数 $f(x)$ 的连续性推知，对任意正数 ε，总可以找到一种分割，把区间 $[a, b]$ 分为 $[x_0, x_1]$，$[x_1, x_2]$，…，$[x_{N-1}, x_N]$（其中 $x_0=a$，$x_N=b$）等 N 个小区间，使得当 $x\in[x_k, x_{k+1}]$ 时，$|f(x)-f(x_k)|<\varepsilon$. 利用这种情况，我们能导入一个辅助函数 $f_\varepsilon(x)$，它只取有限个值，并且当 $x_k<x<x_{k+1}$ 时，$f_\varepsilon(x)=f(x_k)$.

这样显然对 $a\leqslant x\leqslant b$ 的一切 x 皆有不等式

$$|f(x)-f_\varepsilon(x)|<\varepsilon \tag{5.2.9}$$

在此我们可预先选取分点 $x_1,x_2,\cdots,x_{N-1}$，使它们是 $F(x)$ 的连续点. 因为 $\{F_n(x)\}$ 弱收敛于 $F(x)$，故当 n 充分大时，在此 $N-1$ 个分点及 x_0,x_N 上成立不等式

$$|F(x_k)-F_n(x_k)|<\frac{\varepsilon}{MN} \tag{5.2.10}$$

这里 M 是 $|f(x)|$ 在区间 $a\leqslant x\leqslant b$ 中的最大值. 显然,

$$\begin{aligned}
&\left|\int_a^b f(x)\mathrm{d}F(x)-\int_a^b f(x)\mathrm{d}F_n(x)\right| \\
\leqslant&\left|\int_a^b f(x)\mathrm{d}F(x)-\int_a^b f_\varepsilon(x)\mathrm{d}F(x)\right| \\
&+\left|\int_a^b f_\varepsilon(x)\mathrm{d}F(x)-\int_a^b f_\varepsilon(x)\mathrm{d}F_n(x)\right| \\
&+\left|\int_a^b f_\varepsilon(x)\mathrm{d}F_n(x)-\int_a^b f(x)\mathrm{d}F_n(x)\right|
\end{aligned} \tag{5.2.11}$$

由于(5.2.9),

$$\left|\int_a^b f(x)\mathrm{d}F(x)-\int_a^b f_\varepsilon(x)\mathrm{d}F(x)\right|\leqslant\varepsilon[F(b)-F(a)] \tag{5.2.12}$$

$$\begin{aligned}
&\left|\int_a^b f_\varepsilon(x)\mathrm{d}F_n(x)-\int_a^b f(x)\mathrm{d}F_n(x)\right| \\
&\qquad\leqslant\varepsilon[F_n(b)-F_n(a)]
\end{aligned} \tag{5.2.13}$$

而由(5.2.9),(5.2.10)可知

$$\begin{aligned}
&\left|\int_a^b f_\varepsilon(x)\mathrm{d}F(x)-\int_a^b f_\varepsilon(x)\mathrm{d}F_n(x)\right| \\
=&\Big|\sum_{k=0}^{N-1}f(x_k)[F(x_{k+1})-F(x_k)] \\
&-\sum_{k=0}^{N-1}f(x_k)[F_n(x_{k-1})-F_n(x_k)]\Big| \\
=&\Big|\sum_{k=0}^{N-1}f(x_k)[F(x_{k+1})-F_n(x_{k+1})] \\
&-\sum_{k=0}^{N-1}f(x_k)[F(x_k)-F_n(x_k)]\Big| \\
\leqslant&N\left[M\frac{\varepsilon}{MN}+M\frac{\varepsilon}{MN}\right]=2\varepsilon
\end{aligned} \tag{5.2.14}$$

因此

$$\left|\int_a^b f(x)\mathrm{d}F(x)-\int_a^b f(x)\mathrm{d}F_n(x)\right|$$
$$\leqslant\varepsilon[F(b)-F(a)]+\varepsilon[F_n(b)-F_n(a)]+2\varepsilon$$

由于$\{F_n(x)\}$的一致有界性，上式右边可以任意小，故定理得证.

定理 5.2.3（拓广的海莱第二定理） 设$f(x)$在$(-\infty, \infty)$上有界连续，又$\{F_n(x)\}$是$(-\infty, +\infty)$上弱收敛于函数$F(x)$的一致有界非降函数序列，且

$$\lim_{n\to\infty}F_n(-\infty)=F(-\infty),\lim_{n\to\infty}F_n(+\infty)=F(+\infty)$$

则

$$\lim_{n\to\infty}\int_{-\infty}^{\infty} f(x)\mathrm{d}F_n(x)=\int_{-\infty}^{\infty} f(x)\mathrm{d}F(x)$$

［证明］ 设$A<0$，$B>0$，令

$$J_1=\left|\int_{-\infty}^{A} f(x)\mathrm{d}F_n(x)-\int_{-\infty}^{A} f(x)\mathrm{d}F(x)\right|$$
$$J_2=\left|\int_{A}^{B} f(x)\mathrm{d}F_n(x)-\int_{A}^{B} f(x)\mathrm{d}F(x)\right|$$
$$J_3=\left|\int_{B}^{\infty} f(x)\mathrm{d}F_n(x)-\int_{B}^{\infty} f(x)\mathrm{d}F(x)\right|$$

显然

$$\left|\int_{-\infty}^{\infty} f(x)\mathrm{d}F_n(x)-\int_{-\infty}^{\infty} f(x)\mathrm{d}F(x)\right|\leqslant J_1+J_2+J_3$$

由于$f(x)$是有界的，存在常数$M>0$，使$|f(x)|<M$. 又由于序列$\{F_n(x)\}$的一致有界性，只要A与B的绝对值充分大，并使A和B是$F(x)$的连续点，而n也取得充分大，则可使J_1,J_3小到预先给定的程度. 事实上

$$J_1\leqslant M[F_n(A)-F_n(-\infty)+F(A)-F(-\infty)]$$
$$J_3\leqslant M[F_n(+\infty)-F_n(B)+F(+\infty)-F(B)]$$

而按假定有

$$\lim_{n\to\infty}F_n(A)=F(A),\lim_{n\to\infty}F_n(B)=F(B)$$

$$\lim_{n\to\infty}F_n(-\infty)=F(-\infty),\lim_{n\to\infty}F_n(+\infty)=F(+\infty)$$

故 J_1,J_3 可以任意小. 再根据定理 5.2.2,只要 n 充分大,也可使 J_2 任意小,从而证得了定理.

二、连续性定理

下面我们将导出一个分布函数列弱收敛到一个极限分布的充要条件，这个结果同时说明了存在于分布函数与特征函数之间的一一对应是连续的，这个性质对于特征函数成为研究一些极限定理的主要工具有基本的重要性.

定理 5.2.4（正极限定理） 设分布函数列$\{F_n(x)\}$弱收敛于某一分布函数 $F(x)$,则相应的特征函数列$\{f_n(t)\}$收敛于特征函数 $f(t)$,且在 t 的任一有限区间内收敛是一致的.

［证明］ 函数 e^{itx}在$-\infty<x<+\infty$上有界连续,而

$$f_n(t)=\int_{-\infty}^{\infty}e^{itx}dF_n(x)$$

$$f(t)=\int_{-\infty}^{\infty}e^{itx}dF(x)$$

因此由拓广的海莱第二定理即知当 $n\to\infty$时，有

$$f_n(t)\to f(t)$$

至于在 t 的每一有限区间内收敛的一致性，由拓广的海莱第二定理的证明就可看出.

定理 5.2.5（逆极限定理） 设特征函数列 $\{f_n(t)\}$ 收敛于某一函数 $f(t)$，且 $f(t)$ 在 $t=0$ 连续，则相应的分布函数列 $\{F_n(x)\}$ 弱收敛于某一分布函数 $F(x)$，而且 $f(t)$ 是 $F(x)$ 的特征函数.

［证明］ 由海莱第一定理，知必存在子序列 $\{F_{n_k}(x)\}$ 弱收敛于某一非降函数 $F(x)$,且 $F(x)$可视为左连续的. 极限函数 $F(x)$显然满足 $F(-\infty)\geqslant0$,$F(+\infty)\leqslant1$,我们来证明 $F(x)$是分布函数. 否则,应有

$$\delta = F(+\infty) - F(-\infty) < 1 \tag{5.2.15}$$

任取一正数 $\varepsilon < 1-\delta$. 因 $f(t)$是特征函数列的极限，故 $f(0)=1$. 由于 $f(t)$在 $t=0$ 是连续的，故可选取充分小的正数 τ，使

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)\mathrm{d}t\right| > 1-\frac{\varepsilon}{2} > \delta+\frac{\varepsilon}{2} \tag{5.2.16}$$

同时选取 $X \geqslant \frac{4}{\tau\varepsilon}$及 K，使 $k \geqslant K$ 时，

$$\delta_k = F_{n_k}(X) - F_{n_k}(-X) \leqslant \delta + \frac{\varepsilon}{4}$$

又因 $f_{n_k}(t)$是特征函数，那末

$$\int_{-\tau}^{\tau} f_{n_k}(t)\mathrm{d}t = \int_{-\infty}^{\infty}\left[\int_{-\tau}^{\tau} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}t\right]\mathrm{d}F_{n_k}(x) \tag{5.2.17}$$

显然

$$\left|\int_{-\tau}^{\tau} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}t\right| \leqslant 2\tau$$

还有，在 $|x|>X$ 时，

$$\left|\int_{-\tau}^{\tau} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}t\right| = \left|\frac{2}{x}\sin\tau x\right| < \frac{2}{X} \tag{5.2.18}$$

因此

$$\begin{aligned}\left|\int_{-\tau}^{\tau} f_{n_k}(t)\mathrm{d}t\right| &\leqslant \left|\int_{|x|\leqslant X}\left(\int_{-\tau}^{\tau} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}t\right)\mathrm{d}F_{n_k}(x)\right| \\ &\quad + \left|\int_{|x|>X}\left(\int_{-\tau}^{\tau} \mathrm{e}^{\mathrm{i}tx}\mathrm{d}t\right)\mathrm{d}F_{n_k}(x)\right| \\ &< 2\tau\delta_k + \frac{2}{X}\end{aligned}$$

所以

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f_{n_k}(t)\mathrm{d}t\right| \leqslant \delta_k + \frac{1}{X\tau} \leqslant \delta + \frac{\varepsilon}{2}$$

令 $k\to\infty$，由控制收敛定理知

$$\frac{1}{2\tau}\left|\int_{-\tau}^{\tau} f(t)\mathrm{d}t\right| \leqslant \delta + \frac{\varepsilon}{2}$$

这与(5.2.16)矛盾，因此(5.2.15)不成立，也即应有

$$F(-\infty)=0, F(+\infty)=1$$

因而 $F(x)$是分布函数. 再由定理 5.2.4 推知 $f(t)$是 $F(x)$的特征函数.

进而证明$\{F_n(x)\}$也弱收敛于同一分布函数 $F(x)$. 如其不然，一定存在 $F(x)$的一个连续点 x_0 使$\{F_n(x_0)\}$不收敛于 $F(x_0)$. 这时可从$\{F_n(x_0)\}$中选取一个收敛的子序列$\{F_{m_k}(x_0)\}$，其极限 $F^*(x_0)\neq F(x_0)$. 根据海莱第一定理，一定可以选取$\{F_{m_k}(x)\}$的一个子序列$\{F_{m_{k_l}}(x)\}$弱收敛于某一有界的非降函数 $F^*(x)$，这个极限函数至少在 x_0 点与 $F(x)$不相等. 但重复前面的论证可知 $F^*(x)$亦应是分布函数，其对应的特征函数也是 $f(t)$，由唯一性定理，我们又有 $F^*(x)=F(x)$，引出了矛盾. 故$\{F_n(x)\}$弱收敛于 $F(x)$，于是证得定理.

在逆极限定理中，若保留"特征函数列$\{f_n(t)\}$收敛于某一函数 $f(t)$"的要求，而把"$f(t)$在 $t=0$ 连续"的要求改成"特征函数列$\{f_n(t)\}$在包含原点的某一区间中一致收敛于函数 $f(t)$"，则定理的结论仍然成立. 这是因为由一致收敛性及 $f_n(t)$在原点的连续性可以推知 $f(t)$在原点的连续性. 通常把"特征函数列$\{f_n(t)\}$在$(-\infty,\infty)$上的任一有限闭区间中都一致收敛于一个函数 $f(t)$"简称为"$\{f_n(t)\}$内闭匀敛于 $f(t)$"，这样我们就可以把分布函数列$\{F_n(x)\}$弱收敛于某一分布函数的充要条件简述为：它相应的特征函数列$\{f_n(t)\}$内闭匀敛于某一函数 $f(t)$.

通常把正逆极限定理合称**连续性定理**，因为它们表述了分布函数与特征函数一一对应关系的"连续性". 这定理最先由法国数学家莱维(Lévy)及瑞典数学家克拉默(Cramér)证得，因此又称**莱维—克拉默定理**.

三、随机变量的收敛性

概率论中的极限定理研究的是随机变量序列的某种收敛性，

对随机变量收敛性的不同定义将导致不同的极限定理，而随机变量的收敛性的确可以有各种不同的定义，现在就来讨论这个问题.

首先，分布函数弱收敛的讨论启发我们引进如下定义.

定义 5.2.2（依分布收敛） 设随机变量 $\xi_n(\omega)$、$\xi(\omega)$的分布函数分别为 $F_n(x)$及 $F(x)$，如果 $F_n(x)\xrightarrow{W}F(x)$，则称$\{\xi_n(\omega)\}$**依分布收敛**(**convergence in distribution**)于 $\xi(\omega)$，并记为 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$.

其次，由伯努利大数定律，我们很自然地引进下面的定义.

定义 5.2.3（依概率收敛） 如果

$$\lim_{n\to\infty}P\{|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon\}=0 \tag{5.2.19}$$

对任意的 $\varepsilon>0$ 成立，则称$\{\xi_n(\omega)\}$**依概率收敛**(**convergence in probability**)于 $\xi(\omega)$，并记为 $\xi_n(\omega)\xrightarrow{P}\xi(\omega)$.

这样一来，伯努利大数定律可以重新叙述如下：

设 μ_n 是 n 次独立试验中事件 A 出现的次数，而 p 是事件 A 在每次试验中出现的概率，则频率$\frac{\mu_n}{n}$依概率收敛于概率 p.

上述两种收敛性之间的关系可以从下面定理中看到，这也说明了随机变量序列依概率收敛性的重要性.

定理 5.2.6 $\xi_n\xrightarrow{P}\xi\Rightarrow\xi_n\xrightarrow{L}\xi$.

［证明］ 因为，对 $x'<x$ 有

$$\begin{aligned}\{\xi<x'\}&=\{\xi_n<x,\xi<x'\}+\{\xi_n\geqslant x,\xi<x'\}\\&\subset\{\xi_n<x\}+\{\xi_n\geqslant x,\xi<x'\}\end{aligned}$$

所以我们有

$$F(x')\leqslant F_n(x)+P\{\xi_n\geqslant x,\xi<x'\}$$

如果$\{\xi_n\}$依概率收敛于 ξ，则

$$P\{\xi_n\geqslant x,\xi<x'\}\leqslant P\{|\xi_n-\xi|\geqslant x-x'\}\to 0$$

因而有

$$F(x')\leqslant\varliminf_{n\to\infty}F_n(x)$$

同理可证，对 $x''>x$，成立

$$\overline{\lim_{n\to\infty}}F_n(x)\leqslant F(x'')$$

所以对 $x'<x<x''$，有

$$F(x')\leqslant\underline{\lim_{n\to\infty}}F_n(x)\leqslant\overline{\lim_{n\to\infty}}F_n(x)\leqslant F(x'')$$

如果 x 是 $F(x)$的连续点，则令 x'，x''趋于 x 可得

$$F(x)=\lim_{n\to\infty}F_n(x)$$

定理证毕.

由于不同的随机变量可以对应于同一分布函数，因此一般地讲，由分布函数列的收敛性当然推不出随机变量序列的其他收敛性. 试看下例.

［例 3］ 若样本空间 $\Omega=\{\omega_1,\omega_2\}$，$P(\omega_1)=P(\omega_2)=\frac{1}{2}$，定义随机变量 $\xi(\omega)$如下：$\xi(\omega_1)=-1$，$\xi(\omega_2)=1$，则 $\xi(\omega)$的分布列为

$$\begin{pmatrix}-1 & 1\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix} \tag{5.2.20}$$

若对一切 n，令 $\xi_n(\omega)=-\xi(\omega)$，显然 $\xi_n(\omega)$的分布列也是(5.2.20)，因此 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$. 但是对任意的 $0<\varepsilon<2$，

$$P\{|\xi_n(\omega)-\xi(\omega)|>\varepsilon\}=P\{\Omega\}=1$$

因此$\{\xi_n(\omega)\}$不依概率收敛于 $\xi(\omega)$，这为定理 5.2.6 之逆提供了反例.

进一步，若令 $\xi_{2n}(\omega)=\xi(\omega)$，$\xi_{2n+1}(\omega)=-\xi(\omega)$则 $\xi_n(\omega)\xrightarrow{L}\xi(\omega)$依然成立，但此时对随机变量序列$\{\xi_n(\omega)\}$实在很难有其他的收敛性可言.

但是，在特殊场合却有下面结果.

定理 5.2.7 设 C 是常数，则 $\xi_n\xrightarrow{P}C\Leftrightarrow\xi_n\xrightarrow{L}C$.

［证明］ 由定理 5.2.6 可知只须证明由依分布收敛于常数可推出依概率收敛于常数. 事实上，对任意的 $\varepsilon>0$，

$$\begin{aligned}P\{|\xi_n-C|\geqslant\varepsilon\}&=P\{\xi_n\geqslant C+\varepsilon\}+P\{\xi_n\leqslant C-\varepsilon\}\\&=1-F_n(C+\varepsilon)+F_n(C-\varepsilon+0)\end{aligned}$$

$$\to 1-1+0=0$$

$$(n\to\infty)$$

因为有的大数定律是讨论随机变量序列收敛于常数的，这时将用到上述结果.

仔细考察上节关于大数定律的证明，有助于理解下面关于随机变量收敛性的第三种定义.

定义 5.2.4（*r* 阶收敛） 设对随机变量 ξ_n 及 ξ 有 $E|\xi_n|^r<\infty$，$E|\xi|^r<\infty$，其中 $r>0$ 为常数，如果

$$\lim_{n\to\infty} E|\xi_n-\xi|^r=0 \tag{5.2.21}$$

则称 $\{\xi_n\}$ ***r* 阶收敛**（***convergence in r-order mean***）于 ξ，并记为 $\xi_n \xrightarrow{r} \xi$.

下面定理揭示了 r 阶收敛与依概率收敛的关系.

定理 5.2.8 $\xi_n \xrightarrow{r} \xi \Rightarrow \xi_n \xrightarrow{P} \xi$.

［证明］ 先证对于任意 $\varepsilon>0$，成立

$$P\{|\xi_n-\xi|\geqslant\varepsilon\}\leqslant\frac{E|\xi_n-\xi|^r}{\varepsilon^r} \tag{5.2.22}$$

事实上，若以 $F(x)$ 记 $\xi_n-\xi$ 的分布函数，则仿车贝晓夫不等式的证明可得

$$\begin{aligned}P\{|\xi_n-\xi|\geqslant\varepsilon\}&=\int_{|x|\geqslant\varepsilon}\mathrm{d}F(x)\\&\leqslant\int_{|x|\geqslant\varepsilon}\frac{|x|^r}{\varepsilon^r}\mathrm{d}F(x)\leqslant\frac{1}{\varepsilon^r}\int_{-\infty}^{\infty}|x|^r\mathrm{d}F(x)\\&=\frac{E|\xi_n-\xi|^r}{\varepsilon^r}\end{aligned}$$

不等式(5.2.22)是车贝晓夫不等式的推广，通常称作**马尔可夫不等式**，当 $r=2$ 时就是车贝晓夫不等式. 定理 5.2.8 是马尔可夫不等式的直接推论.

下例说明定理 5.2.8 之逆不真.

［例 4］ 取 $\Omega=(0,1]$，$\mathscr{F}$ 为(0,1]中博雷尔点集全体所构成

的 σ 域，P 为勒贝格测度. 定义 $\xi(\omega)\equiv 0$ 及

$$\xi_n(\omega)=\begin{cases} n^{1/r}, 0<\omega\leqslant\dfrac{1}{n} \\ 0, \quad \dfrac{1}{n}<\omega\leqslant 1 \end{cases} \tag{5.2.23}$$

显然对一切 $\omega\in\Omega$，$\xi_n(\omega)\to\xi(\omega)$，又对于任意的 $\varepsilon>0$，

$$P\{|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon\}\leqslant\frac{1}{n}$$

因此 $\xi_n\xrightarrow{P}\xi$，但是

$$E|\xi_n-\xi|^r=(n^{1/r})^r\cdot\frac{1}{n}=1$$

在 r 阶收敛中，最重要的是 $r=2$ 的情况，这时称为**均方收敛**.

下面是关于随机变量收敛性的第四种定义.

定义 5.2.5（以概率 1 收敛） 如果

$$P\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1 \tag{5.2.24}$$

则称 $\{\xi_n(\omega)\}$ **以概率 1（convergence in probability 1）收敛于** $\xi(\omega)$，又称 $\{\xi_n(\omega)\}$ **几乎处处收敛于** $\xi(\omega)$，记为 $\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega)$.

以概率 1 收敛是概率论中较强的一种收敛性，但是正如例 4 所表明的，一般并不能由它推出 r 阶收敛. 关于以概率 1 收敛的讨论将在 § 4 中继续进行，在那里将证明可由以概率 1 收敛推出依概率收敛.

*四、波赫纳尔（Bochner）—辛钦（Xинчин）定理

利用这个机会，我们来叙述并证明一个关于特征函数的重要定理.

定理 5.2.9（波赫纳尔—辛钦） 函数 $f(t)$ 是特征函数的充要条件是：$f(t)$ 非负定，连续，且 $f(0)=1$.

在证明的过程中，顺带证明了一个在随机过程中将用到的与上述定理类似的赫尔格洛兹（Herglotz）定理. 为此，有

定义 5.2.6 如果对任意的正整数 n 及复数 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 均有

$$\sum_{k=1}^{n}\sum_{j=1}^{n}C_{k-j}\lambda_k\bar{\lambda}_j\geqslant 0 \qquad (5.2.25)$$

则称复数列 $C_n(n=0,\pm 1,\pm 2,\cdots)$是**非负定的**.

定理 5.2.10(赫尔格洛兹) 数列 $C_n(n=0,\pm 1,\pm 2,\cdots)$可以表为

$$C_n=\int_{-\pi}^{\pi}e^{inx}dG(x) \qquad (5.2.26)$$

的充要条件是它是非负定的,其中 $G(x)$是$[-\pi,\pi]$上有界、非降、左连续函数.

定理 5.2.9 的必要性已在第四章 §5 中证过,定理 5.2.10 也可类似证明,下面只须证明充分性.

由于 $f(t)$是非负定的,故对任何 N,实数$\dfrac{k}{n}$及复数 $e^{-ikx}(k=0,1,\cdots,N-1)$,皆有

$$\mathscr{P}_N^{(n)}(x)=\frac{1}{N}\sum_{k=0}^{N-1}\sum_{j=0}^{N-1}f\left(\frac{k-j}{n}\right)e^{-i(k-j)x}\geqslant 0$$

易知,其中使 $k-j$ 等于 r 的项有 $N-|r|$个,r 可由$-N+1$ 变到 $N-1$. 因此

$$\mathscr{P}_N^{(n)}(x)=\sum_{r=-N}^{N}\left(1-\frac{|r|}{N}\right)f\left(\frac{r}{n}\right)e^{-irx}$$

从而

$$\begin{aligned}&\int_{-\pi}^{\pi}e^{isx}\mathscr{P}_N^{(n)}(x)dx\\&=\sum_{r=-N}^{N}\left(1-\frac{|r|}{N}\right)f\left(\frac{r}{n}\right)\int_{-\pi}^{\pi}e^{-irx}e^{isx}dx\end{aligned}$$

由于

$$\int_{-\pi}^{\pi}e^{-i(r-s)x}dx=\begin{cases}0, & r\neq s\\ 2\pi, & r=s\end{cases}$$

所以

$$\left(1-\frac{|s|}{N}\right)f\left(\frac{s}{n}\right)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}sx}\mathscr{P}_N^{(n)}(x)\mathrm{d}x$$

$$=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}sx}\mathrm{d}F_N^{(n)}(x)$$

其中

$$F_N^{(n)}(x)=\frac{1}{2\pi}\int_{-\pi}^{x}\mathscr{P}_N^{(n)}(t)\mathrm{d}t$$

是一个在$[-\pi,\pi]$上有界的非降函数，其全变差为

$$F_N^{(n)}(\pi)=\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathscr{P}_N^{(n)}(t)\mathrm{d}t=f(0)=1$$

补充定义$x<-\pi$时$F_N^{(n)}(x)=0$，$x>\pi$时，$F_N^{(n)}(x)=1$，则$F_N^{(n)}(x)$是一分布函数.

按海莱第一定理，存在序列N_k，$k\to\infty$时$N_k\to\infty$，使函数序列$F_{N_k}^{(n)}(x)$弱收敛于某一非降函数$F^{(n)}(x)$. 又因对任何N及$\varepsilon>0$，

$$F_N^{(n)}(-\pi-\varepsilon)=0,\quad F_N^{(n)}(\pi+\varepsilon)=1$$

因而也有

$$F^{(n)}(-\pi-\varepsilon)=0,\quad F^{(n)}(\pi+\varepsilon)=1$$

所以$F^{(n)}(x)$也是分布函数.

按海莱第二定理，

$$\lim_{k\to\infty}\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}sx}\mathrm{d}F_{N_k}^{(n)}(x)=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}sx}\mathrm{d}F^{(n)}(x)$$

所以，对一切整数s（$s=0$，±1，±2，…）有

$$f\left(\frac{s}{n}\right)=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}sx}\mathrm{d}F^{(n)}(x)$$

至此我们已顺便证明了赫尔格洛兹定理. 特别地

$$f\left(\frac{1}{n}\right)=\int_{-\pi}^{\pi}\mathrm{e}^{\mathrm{i}x}\mathrm{d}F^{(n)}(x)\qquad(5.2.27)$$

考虑特征函数序列

$$f_n(t)=\int_{-n\pi}^{n\pi}\mathrm{e}^{\mathrm{i}tx}\mathrm{d}F_n(x)$$

其中 $F_n(x)=F^{(n)}\left(\frac{x}{n}\right)$. 易知对一切整数 k 有

$$f_n\left(\frac{k}{n}\right)=f\left(\frac{k}{n}\right) \tag{5.2.28}$$

对任何 t，我们总能选取序列 $k=k(n,t)$，使 $0\leqslant t-\frac{k}{n}<\frac{1}{n}$. 由于 $f(t)$连续，从而

$$f(t)=\lim_{n\to\infty}f\left(\frac{k}{n}\right)=\lim_{n\to\infty}f_n\left(\frac{k}{n}\right) \tag{5.2.29}$$

如能证明对一切实数 t，有

$$f(t)=\lim_{n\to\infty}f_n(t) \tag{5.2.30}$$

那么由逆极限定理即知 $f(t)$是特征函数了.

为此，由(5.2.28)和(5.2.29)有

$$\begin{aligned}\lim_{n\to\infty}f_n(t)&=\lim_{n\to\infty}\left\{\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]+f_n\left(\frac{k}{n}\right)\right\}\\&=f(t)+\lim_{n\to\infty}\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]\end{aligned} \tag{5.2.31}$$

令 $\theta=t-\frac{k}{n}$，那末 $0\leqslant\theta<\frac{1}{n}$. 按 $f_n(t)$的定义，有

$$\begin{aligned}\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|&=\left|\int_{-n\pi}^{n\pi}e^{i(k/n)x}(e^{i\theta x}-1)dF_n(x)\right|\\&\leqslant\int_{-n\pi}^{n\pi}|e^{i\theta x}-1|dF_n(x)\end{aligned} \tag{5.2.32}$$

利用施瓦茨不等式，可得

$$\begin{aligned}\int_{-n\pi}^{n\pi}|e^{i\theta x}-1|dF_n(x)&\leqslant\sqrt{\int_{-n\pi}^{n\pi}|e^{i\theta x}-1|^2dF_n(x)}\\&=\left[\int_{-n\pi}^{n\pi}2(1-\cos\theta x)dF_n(x)\right]^{\frac{1}{2}}\\&=[2(1-\mathrm{Re}f_n(\theta))]^{\frac{1}{2}}\end{aligned} \tag{5.2.33}$$

其中 $\mathrm{Re}f_n(\theta)$为 $f_n(\theta)$的实数部分. 既然在 $0\leqslant\alpha<1$ 及 $-\pi\leqslant y<\pi$ 时有 $\cos y\leqslant\cos\alpha y$，则

$$1-\mathrm{Re}f_n(\theta)=\int_{-n\pi}^{n\pi}(1-\cos\theta x)\mathrm{d}F_n(x)$$

$$=\int_{-\pi}^{\pi}(1-\cos\theta ny)\mathrm{d}F_n(ny)\leqslant\int_{-\pi}^{\pi}(1-\cos y)\mathrm{d}F_n(ny)$$

$$=\int_{-\pi}^{\pi}(1-\cos y)\mathrm{d}F^{(n)}(y)$$

$$=1-\mathrm{Re}\int_{-\pi}^{\pi}e^{iy}\mathrm{d}F^{(n)}(y)$$

再由(5.2.27)我们得到

$$1-\mathrm{Re}f_n(\theta)\leqslant 1-\mathrm{Re}f\left(\frac{1}{n}\right) \qquad (5.2.34)$$

合并(5.2.31),(5.2.32),(5.2.33),即得

$$\left|f_n(t)-f_n\left(\frac{k}{n}\right)\right|\leqslant\sqrt{2\left(1-\mathrm{Re}f\left(\frac{1}{n}\right)\right)}$$

注意到 $f(0)=1$,则由 $f(t)$的连续性推得

$$\lim_{n\to\infty}\left[f_n(t)-f_n\left(\frac{k}{n}\right)\right]=0$$

于是由(5.2.31)知(5.2.30)成立,定理证毕.

若不证赫尔格洛兹定理而直接证波赫纳尔—辛钦定理,有比较简练的证法,可参看 C. R. Rao, Linear Statistical Inference and its Applications, second edition, pp. 141—142.

§3. 独立同分布场合的极限定理

一、独立和问题

在§1中,我们讨论了伯努利试验场合事件 A 出现次数 μ_n 的极限行为,曾指出 μ_n 可以表示为 n 个独立随机变量之和(以后简称"独立和"),并对它证明了大数定律及中心极限定理,后来又看到这些定理有重要应用.这里自然会提出这样一个问题:这些性质是否只在伯努利试验场合才具有?

研究表明，许多独立和具有类似的性质，本节就要进一步讨论这个问题.

独立和的问题经常出现，例如测量一物体的某种尺寸，如测量一个圆柱体的直径 d，通常采用的办法是对它进行 n 次测量，得到数值 $\xi_1,\xi_2,\cdots,\xi_n$，然后采用平均值

$$\eta_n=\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}$$

作为 d 的数值. 我们知道，测量时有各种随机因素影响，因此其结果带有随机性，这时 η_n 是随机变量之和，如果各次测量是独立的，η_n 便是独立和. 为了说明上面所用办法的合理性就必须研究独立和. 但这里的 ξ_n 不服从伯努利 0—1 分布，因此已不是伯努利试验场合的问题了.

在数理统计中已经把上述做法一般化. 为了研究母体（它通常描述我们感兴趣的某一类现象）的某些特征，就对母体进行若干次观察以得到一批观察值 $\xi_1,\xi_2,\cdots,\xi_n$，并称它们是一个容量为 n 的子样. 再利用这个子样来构造各种统计量，例如

$$\frac{\xi_1+\xi_2+\cdots+\xi_n}{n}\quad 或 \quad \frac{\xi_1^2+\xi_2^2+\cdots+\xi_n^2}{n}$$

以对母体的特征作各种推断. 虽然每次观察得到的是具体的数值，但是为了比较各个统计量或各种推断方法的优劣，有必要把这些观察看作是某随机变量 ξ 的观察值，通常假定 $\xi_1,\xi_2,\cdots,\xi_n$ 是相互独立的，且它们与 ξ 具有相同的概率分布，这时上面的两个统计量便都是独立同分布（independent and identically distribution）的随机变量之和.

独立和的问题在许多实际问题中也出现，例如在计算电车整流站的电力负荷时，就遇到独立和问题，因为整流站的电力负荷等于各电车使用电力之和，每辆电车在某时刻的用电量是随机的，作为初步近似，可以假定各电车的用电量是相互独立的，因此这里遇到的正是独立和. 在车间用电问题中，若有多类车床，用电量各不

相同，则总用电量也需通过独立和来计算.

可见独立和问题经常遇到，而且各加项一般都不是 0—1 分布. 本节专门讨论各个加项服从相同分布的场合，这是实际工作（特别是在数理统计）中最常碰到的. 从数学方面来看，这是最简单、最基本、最便于处理的，而且所用的处理方法可以相当方便地运用到更一般的场合.

在上章§5中，我们讨论过独立的服从同类型分布的随机变量之和的分布问题，证明了某些分布的再生性，使用的工具是特征函数. 现在我们要处理的问题与那里有很大区别，首先我们将对很一般的分布进行讨论，因而再生性通常都不满足；其次，不是对固定的 n 进行讨论，而是讨论 $n\to\infty$ 时的情况，即研究极限定理. 从数学的角度来看，它们可以看作是伯努利试验场合极限定理的推广，这里也研究大数定律与中心极限定理.

所使用的工具还是特征函数. 我们已经看到它很适合于处理独立和问题，有了上节的连续性定理，我们将进一步看到，它也很适合于处理极限分布问题. 事实上，正是由于特征函数这一有力工具的使用，使得所有古典极限定理在短期内便得到了完满的解决. 拉普拉斯就已经知道并应用了特征函数，俄国数学家李雅普诺夫（Ляпунов）最先发现并证明了收敛于正态分布的连续性定理（但并未明确叙述），从那时起，特征函数的理论不断得到完善. 在这当中，法国数学家莱维有突出的贡献. 现在，特征函数法已经成了概率论的基本方法之一.

二、辛钦大数定律

在§1中，我们已经通过车贝晓夫不等式建立起多种大数定律，那里都假定了方差的存在性，但是在独立同分布场合，并不需要有这个要求，这就是有名的辛钦大数定律告诉我们的. 用特征函数作为工具，这个定理很容易证明.

定理 5.3.1（辛钦） 设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$是相互独立的随机变

量序列，它们服从相同的分布，且具有有限的数学期望

$$a = E\xi_n$$

则对任意的 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| < \varepsilon\right\} = 1 \qquad (5.3.1)$$

［证明］ 由于 $\xi_1,\xi_2,\cdots,\xi_n$ 具有相同分布，故有同一特征函数，设为 $f(t)$，因为数学期望存在，故 $f(t)$ 可展开成

$$f(t) = f(0) + f'(0)t + o(t) = 1 + \mathrm{i}at + o(t) \qquad (5.3.2)$$

而 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 的特征函数为

$$\left[f\left(\frac{t}{n}\right)\right]^n = \left[1 + \mathrm{i}a\frac{t}{n} + o\left(\frac{t}{n}\right)\right]^n \qquad (5.4.3)$$

对于固定的 t

$$\left[f\left(\frac{t}{n}\right)\right]^n \to \mathrm{e}^{\mathrm{i}at} \quad (n\to\infty) \qquad (5.3.4)$$

极限函数 $\mathrm{e}^{\mathrm{i}at}$ 是连续函数，它是退化分布 $I(x-a)$ 所对应的特征函数，由逆极限定理知 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 的分布函数弱收敛于 $I(x-a)$，再由定理 5.2.7 知 $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ 依概率收敛于常数 a，从而证明了定理.

显然，伯努利大数定律是辛钦大数定律的特殊情况. 辛钦大数定律在应用中很重要，下面通过一个例子来说明.

［例 1］ （用蒙特卡洛方法计算定积分）为计算积分

$$J = \int_a^b g(x)\mathrm{d}x \qquad (5.3.5)$$

可以通过下面概率论方法实现.

任取一列相互独立的、都具有 $[a,b]$ 中均匀分布的随机变量 $\{\xi_i\}$，则 $\{g(\xi_i)\}$ 也是一列相互独立相同分布的随机变量，而且

$$Eg(\xi_i) = \frac{1}{b-a}\int_a^b g(x)\mathrm{d}x = \frac{J}{b-a} \qquad (5.3.6)$$

既然

$$J=(b-a)\cdot Eg(\xi_i) \tag{5.3.7}$$

因此只要能求得 $Eg(\xi_i)$，便能得到 J 的数值.

为求 $Eg(\xi_i)$，自然想到大数定律，因为

$$\frac{g(\xi_1)+g(\xi_2)+\cdots+g(\xi_n)}{n}\xrightarrow{P}Eg(\xi_i) \tag{5.3.8}$$

这样一来，只要能生成随机变量序列 $\{g(\xi_i)\}$ 就能对积分(5.3.5)进行数值计算，而生成 $\{g(\xi_i)\}$ 的关键是要生成相互独立相同分布的 $\{\xi_i\}$，这里的 ξ_i 均服从 $[a,b]$ 上均匀分布.

现在已经可以把上述想法变成现实.这就是在电子计算机上产生服从均匀分布 $[a,b]$ 的随机数 $\{\xi_i\}$ 并利用(5.3.7)(5.3.8)估算 J，这种做法与我们在蒲丰投针问题中通过投针计算圆周率 π 的做法是一致的.这种通过概率论的想法构造模型从而实现数值计算的方法，随着电子计算机的发展，已形成一种新的计算方法——**概率计算方法**，亦称**蒙特卡洛方法**，它在原子物理、公用事业理论中发挥了不少作用，这个方法的理论根据之一就是大数定律.

至于计算积分，蒙特卡洛方法的实用场合是计算重积分

$$I=\int_K g(P)\mathrm{d}P \tag{5.3.9}$$

其中 P 是 m 维空间的点，当 m 较大时，用蒙特卡洛方法比一般数值法有优点，主要是它的误差与维数 m 无关.

三、中心极限定理

我们转而考虑如何把积分极限定理推广到相互独立相同分布，但分布函数为任意的随机变量序列的场合，这类问题在实际应用中非常重要.

若 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是一串相互独立相同分布的随机变量序列，且

$$E\xi_k=m,D\xi_k=\sigma^2 \tag{5.3.10}$$

我们来讨论标准化随机变量和

$$\zeta_n=\frac{1}{\sigma\sqrt{n}}\sum_{k=1}^{n}(\xi_k-m) \tag{5.3.11}$$

的极限分布.

林德贝格(Lindeberg)与莱维建立了下列中心极限定理.

定理 5.3.2(林德贝格—莱维) 对于标准化和(5.3.11),若 $0<\sigma^2<\infty$,则

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t \tag{5.3.12}$$

[证明] 记 ξ_k-m 的特征函数为 $\varphi(t)$,则 ξ_n 的特征函数为 $\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n$. 由于 $E\xi_k=m, D\xi_k=\sigma^2$ 故 $\varphi'(0)=0, \varphi''(0)=-\sigma^2$. 因此

$$\varphi(t)=1-\frac{1}{2}\sigma^2t^2+o(t^2) \tag{5.3.13}$$

所以

$$\left[\varphi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n=\left[1-\frac{1}{2n}t^2+o\left(\frac{t^2}{n}\right)\right]^n\to\mathrm{e}^{-t^2/2} \tag{5.3.14}$$

由于 $\mathrm{e}^{-t^2/2}$是连续函数,它对应的分布函数为 $N(0,1)$,因此由逆极限定理知

$$P\{\zeta_n<x\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t$$

定理证毕.

用这个定理立即可以推出棣莫弗—拉普拉斯积分极限定理.

林德贝格—莱维定理有广泛应用. 在实际工作中,只要 n 足够大,便可以把独立同分布的随机变量之和当作是正态变量. 这种做法在数理统计中用得很普遍,当处理大子样(容量很大的子样)时,

林德贝格—莱维定理是重要工具.

下面我们介绍另外的两个例子.

[例 2] (正态随机数的产生) 在蒙特卡洛方法中经常需要产生服从正态分布的随机数,但是一般计算机只备有产生[0,1]均匀分布随机数(实际上是伪随机数)的程序.怎样通过[0,1]均匀分布的随机数来产生正态随机数呢?这有多种途径,最常用的是利用上述定理来实现.

设 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 是相互独立、均服从[0,1]均匀分布的随机变量,这时定理 5.3.2 的条件得到满足,故 $\xi_1+\xi_2+\cdots+\xi_n$ 渐近于正态变量.一般 n 取不太大的值就可满足实际要求.图 28 中给出了 $n=1,2,3$ 时的图象.在蒙特卡洛方法中,一般取 $n=12$,并用(5.3.15)式得到新的随机数序列.

$$p_1(x)=1,\qquad 0\leqslant x<1$$

$$p_2(x)=\begin{cases}x, & 0\leqslant x<1\\ 2-x, & 1\leqslant x<2\end{cases}$$

$$p_3(x)=\begin{cases}\frac{1}{2}x^2, & 0\leqslant x<1\\ \frac{1}{2}(-2x^2+6x-3), & 1\leqslant x<2\\ \frac{1}{2}(3-x)^2, & 2\leqslant x<3\end{cases}$$

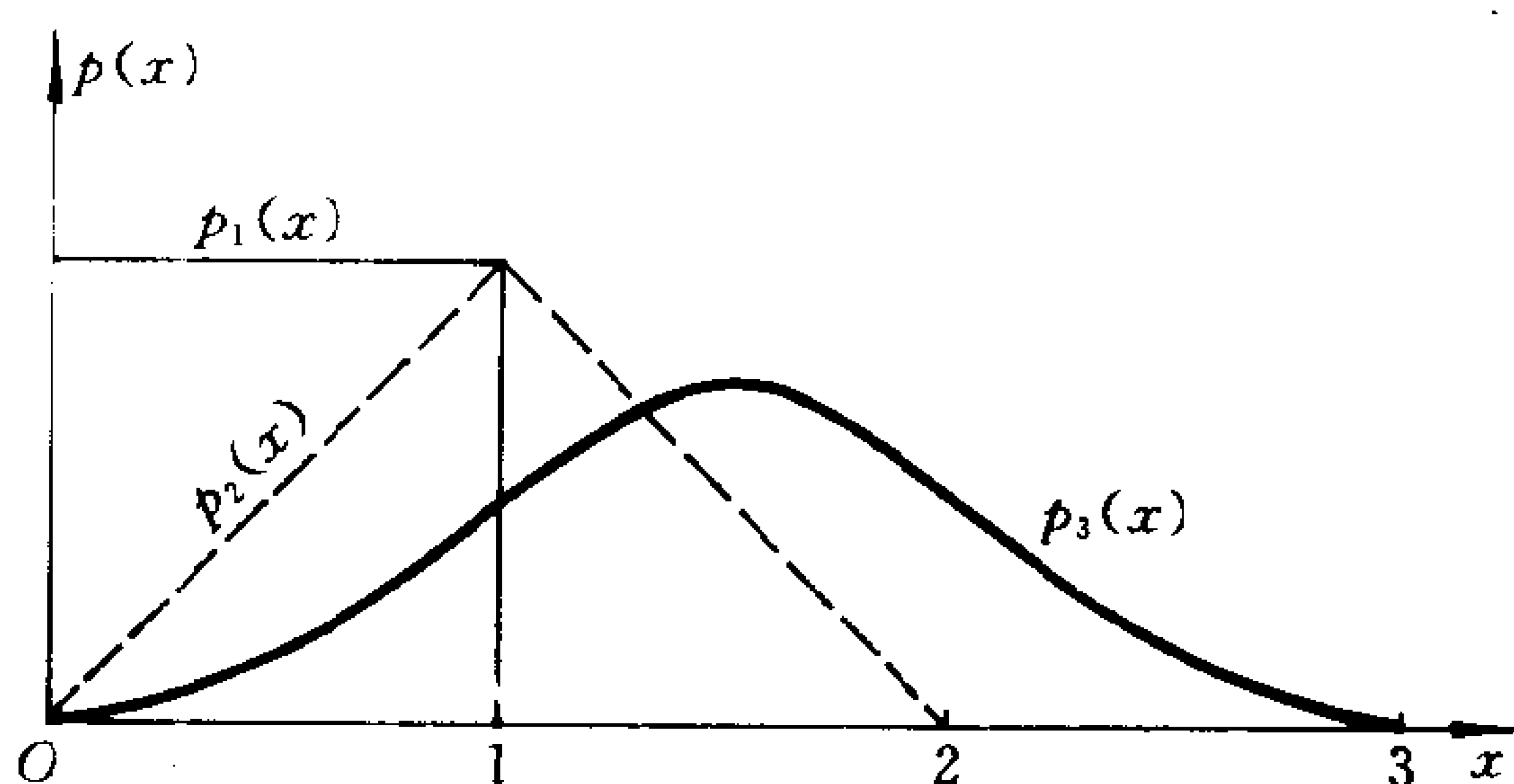


图 28 均匀分布褶积

$$\eta_k=\sum_{i=1}^{12}\xi_{12(k-1)+i}-6,\ k=1,2,\cdots \tag{5.3.15}$$

显然$\{\eta_k\}$也是独立随机数序列，而且$E\eta_k=0, D\eta_k=1$. 经过检验证明，这时η_k的渐近正态性已能满足一般精度要求.

[例 3] （近似数定点运算的误差分析） 数值计算时，任何数x都只能用一定位数的有限小数y来近似，这就产生了一个误差$\xi=x-y$，在下面讨论中，我们假定参加运算的数都用十进制定点表示，每个数都用四舍五入的方法取到小数点后五位，这时相应的舍入误差可以看作是$[-0.5\times10^{-5}, 0.5\times10^{-5})$上的均匀分布.

现在如果要求n个数$x_i(i=1,2,\cdots,n)$的和S，在数值计算中就只能求出相应的有限位小数$y_i(i=1,2,\cdots,n)$的和T，并用T作为S的近似值. 自然要问，这样做造成的误差$\eta=S-T$是多少？

因为我们有

$$S=\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}(y_i+\xi_i)=\sum_{i=1}^{n}y_i+\sum_{i=1}^{n}\xi_i$$

故

$$\eta=\sum_{i=1}^{n}\xi_i$$

一种传统的估计方法是这样的：由于

$$|\xi_i|\leqslant 0.5\times10^{-5}$$

所以

$$|\eta|\leqslant\sum_{i=1}^{n}|\xi_i|\leqslant n\times0.5\times10^{-5}$$

以$n=10000$为例，所得的误差估计为

$$|\eta|\leqslant 0.05 \tag{5.3.16}$$

这种估计方法显然太保守，看来用概率论方法估计是适宜的. 这时直接求$\eta=\sum_{i=1}^{n}\xi_i$的分布不容易，但当n较大时用极限定理作为工具，则能使问题很快得到解决. 因为

$$m=E\xi_i=0,\quad \sigma=\sqrt{D\xi_i}=\frac{0.5\times10^{-5}}{\sqrt{3}}$$

如果假定舍入误差ξ_i是相互独立的，n又较大，那么用定理 5.3.2

得到

$$P\left\{\left|\sum_{i=1}^{n}\xi_i\right|<k\sqrt{n}\,\sigma\right\}\approx\frac{1}{\sqrt{2\pi}}\int_{-k}^{k}e^{-t^2/2}dt$$

取 $k=3$ 时，上式右边为 0.997，因此我们能以 99.7% 的概率断言：

$$|\eta|<3\times100\times\frac{0.5\times10^{-5}}{\sqrt{3}}=0.866\times10^{-3} \tag{5.3.17}$$

这只及(5.3.16)式中上限估计的 60 分之一.

历史上，误差分析是概率论的重要生长点之一. 19 世纪初德国数学家高斯(Gauss)正是在研究测量误差时引进了正态分布并发展了有广泛应用的最小二乘法. 至今这仍是概率论与生产实际有广泛联系的领域之一. 由于高斯对正态分布的研究，因此有不少人称正态分布为**高斯分布**(虽然棣莫弗及拉普拉斯早在高斯之前就把正态分布引入概率论).

下面我们把中心极限定理推广到多变量的场合.

* **定理 5.3.3(多元中心极限定理)** 若 p 维随机向量 $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots, \boldsymbol{\xi}_n, \cdots$ 相互独立，具有相同的分布，其数学期望为 $\boldsymbol{\mu}$，协方差阵为 $\boldsymbol{\Sigma}$，则

$$\boldsymbol{\eta}_n=\{(\boldsymbol{\xi}_1-\boldsymbol{\mu})+(\boldsymbol{\xi}_2-\boldsymbol{\mu})+\cdots+(\boldsymbol{\xi}_n-\boldsymbol{\mu})\}/\sqrt{n} \tag{5.3.18}$$

的极限分布为 $N(\mathbf{0},\boldsymbol{\Sigma})$.

［证明］ 对 p 维列向量 $\boldsymbol{\lambda}$，构造

$$\zeta_n=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\boldsymbol{\lambda}'(\boldsymbol{\xi}_i-\boldsymbol{\mu}) \tag{5.3.19}$$

它是均值为 0，方差为 $\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}$ 的一维随机变量，由定理 5.3.2 知它的分布函数收敛于 $N(0,\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda})$，因此，若以 $f_n(t)$ 记 ζ_n 的特征函数，则

$$f_n(t)\to f(t,\boldsymbol{\lambda})=\exp(-\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}t^2/2)\quad(n\to\infty) \tag{5.3.20}$$

而

$$f_n(t)=Ee^{it\zeta_n}=E\{\exp(it\boldsymbol{\lambda}'\boldsymbol{\eta}_n)\}$$

因而

$$f_n(1)=Ee^{i\zeta_n}=E\{\exp(i\boldsymbol{\lambda}'\boldsymbol{\eta}_n)\}$$

它作为 $\boldsymbol{\lambda}$ 的函数，是 $\boldsymbol{\eta}_n$ 的特征函数. 在(5.3.20)式中，令 $t=1$，得到

$$f_n(1)\to f(1,\boldsymbol{\lambda})=\exp(-\boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}/2) \qquad (5.3.21)$$

这正是 p 维正态分布 $N(\mathbf{0},\boldsymbol{\Sigma})$ 的特征函数. 因此由多维的连续性定理即得结论.

[例 4] 服从(3.2.6)中多项分布的随机向量，可以看作 n 个相互独立相同分布随机向量之和，由定理 5.3.3 可知多项分布渐近于正态分布，真正维数为 $r-1$.

*§ 4. 强大数定律

一、以概率 1 收敛

以前，我们曾顺便提起过以概率 1 收敛及强大数定律，本节将对它们进行深入讨论.

要彻底搞清以概率 1 收敛这个概念，必须对事件(点集)序列的运算有进一步了解，我们就从讨论这个问题开始.

设 $A_1,A_2,\cdots,A_n,\cdots$ 是一列事件，则 $\bigcup\limits_{n=k}^{\infty}A_n$ 表示事件序列 A_k，A_{k+1}，…中至少发生一个，而 $\bigcap\limits_{n=k}^{\infty}A_n$ 则表示 A_k,A_{k+1}，…同时发生. 记

$$\overline{\lim_{n\to\infty}}A_n=\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n \qquad (5.4.1)$$

$$\underline{\lim_{n\to\infty}}A_n=\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}A_n \qquad (5.4.2)$$

称 $\overline{\lim\limits_{n\to\infty}}A_n$ 为事件序列 $\{A_n\}$ 的**上限事件**，它表示 A_n 发生无穷多次，因为 $\omega\in\bigcap\limits_{k=1}^{\infty}\bigcup\limits_{n=k}^{\infty}A_n$ 当且仅当 ω 属于无穷多个 A_n；类似地称 $\underline{\lim\limits_{n\to\infty}}A_n$ 为事件序列 $\{A_n\}$ 的**下限事件**，它表示 A_n 至多只有有限个不发生，因

为 $\omega\in\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}A_n$，当且仅当存在一个 N，使 $\omega\in\bigcap_{n=N}^{\infty}A_n$，因此若 ω 发生，则 $A_N, A_{N+1},\cdots$ 同时发生，这时至多只有前面 $N-1$ 个事件 $A_1, A_2,\cdots, A_{N-1}$ 可能不发生（也可能有些发生）.

显然

$$\overline{\lim_{n\to\infty}}A_n \supset \underline{\lim_{n\to\infty}}A_n \tag{5.4.3}$$

特别当 $\overline{\lim\limits_{n\to\infty}}A_n=\underline{\lim\limits_{n\to\infty}}A_n$ 时，记 $\lim\limits_{n\to\infty}A_n\equiv\overline{\lim\limits_{n\to\infty}}A_n=\underline{\lim\limits_{n\to\infty}}A_n$，并称它为事件序列 $\{A_n\}$ 的**极限事件**.

利用德莫根定理，有

$$\overline{\left(\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\right)}=\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\overline{A}_n$$

$$\overline{\left(\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}A_n\right)}=\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\overline{A}_n$$

因此

$$\underline{\lim_{n\to\infty}}\overline{A}_n=\overline{\left(\overline{\lim_{n\to\infty}}A_n\right)} \tag{5.4.4}$$

$$\overline{\lim_{n\to\infty}}\ \overline{A}_n=\overline{\left(\underline{\lim_{n\to\infty}}A_n\right)} \tag{5.4.5}$$

下面博雷尔—康特立(Cantelli)引理在概率论中有众多的应用.

引理 5.4.1(博雷尔—康特立引理)

(i) 若随机事件序列 $\{A_n\}$ 满足

$$\sum_{n=1}^{\infty}P(A_n)<\infty \tag{5.4.6}$$

则

$$P\{\overline{\lim_{n\to\infty}}A_n\}=0,\qquad P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=1 \tag{5.4.7}$$

(ii) 若 $\{A_n\}$ 是相互独立的随机事件序列，则

$$\sum_{n=1}^{\infty}P(A_n)=\infty \tag{5.4.8}$$

成立的充要条件为

$$P\{\overline{\lim_{n\to\infty}}A_n\}=1 \quad 或 \quad P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=0 \tag{5.4.9}$$

[证明] (i)由于

$$P\{\overline{\lim_{n\to\infty}}A_n\}=P\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n\}$$

$$\leqslant P\{\bigcup_{n=k}^{\infty}A_n\}\leqslant\sum_{n=k}^{\infty}P\{A_n\}\to 0 \qquad (k\to\infty)$$

由(5.4.4)

$$P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=1$$

(ii) 先证必要性. 注意到$\{A_n\}$的独立性,有

$$P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=P\{\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\overline{A}_n\}\leqslant\sum_{k=1}^{\infty}P\{\bigcap_{n=k}^{\infty}\overline{A}_n\}$$

$$=\sum_{k=1}^{\infty}\prod_{n=k}^{\infty}P(\overline{A}_n)=\sum_{k=1}^{\infty}\prod_{n=k}^{\infty}[1-P(A_n)] \tag{5.4.10}$$

由于

$$0\leqslant 1-P(A_n)\leqslant\exp\{-P(A_n)\}$$

则从

$$\sum_{n=1}^{\infty}P(A_n)=\infty$$

可得

$$\prod_{n=k}^{\infty}[1-P(A_n)]\leqslant\lim_{N\to\infty}\exp\left\{-\sum_{n=k}^{N}P(A_n)\right\}=0$$

所以

$$P\{\underline{\lim_{n\to\infty}}\overline{A}_n\}=0$$

再证充分性. 若 $P\{\overline{\lim_{n\to\infty}}A_n\}=1$. 假定$\sum_{n=1}^{\infty}P(A_n)<\infty$,则由(i)得到 $P\{\overline{\lim_{n\to\infty}}A_n\}=0$,产生矛盾. 因 $P(A_n)\geqslant 0$,故只可能是$\sum_{n=1}^{\infty}P(A_n)=\infty$,引理证毕.

现在讨论随机变量序列的以概率 1 收敛性.

若 $\xi_n(\omega)(n=1,2,\cdots)$，$\xi(\omega)$是随机变量，则

$$\{\omega:\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}$$

$$=\left\{\omega:\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)\right\} \qquad (5.4.11)$$

这个式子可以这样理解：因为 $\omega\in\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}$的充要条件是：对任一正整数 m，存在一个正整数 N，使当 $n>N$ 时均有$|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}$；即对任一正整数 m，ω 属于$\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)$的下限事件，这正是(5.4.11)的右边. 从这个表达式中还可以看出，$\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}$是事件，因此

$$P\{\lim_{n\to\infty}\xi_n(\omega)=\xi(\omega)\}=1 \qquad (5.4.12)$$

有明确的意义，这时称$\{\xi_n(\omega)\}$**以概率 1 收敛于** $\xi(\omega)$. 记为$\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega)$.

因此下面两个式子都表达了$\{\xi_n(\omega)\}$以概率 1 收敛于 $\xi(\omega)$.

$$P\left\{\bigcap_{m=1}^{\infty}\bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|<\frac{1}{m}\right)\right\}=1 \qquad (5.4.13)$$

$$P\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}=0 \qquad (5.4.14)$$

进一步，我们要说明，$\{\zeta_n(\omega)\}$以概率 1 收敛于 $\xi(\omega)$的定义也可以表达为：对任意的 $\varepsilon>0$，成立

$$P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon\right\}=0 \qquad (5.4.15)$$

若以 A_n 记$(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)$，上式表示$\overline{\lim_{n\to\infty}}A_n$ 的概率为 0，这与我们对以概率 1 收敛的理解一致. 不过，对于这个结论，还是给它一个严格的证明.

事实上，由于对 $\varepsilon>0$，总有

$$\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}$$

$$\subset\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}$$

因此由(5.4.14)可以推得(5.4.15).反之,利用

$$P\left\{\bigcup_{m=1}^{\infty}\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}$$
$$\leqslant\sum_{m=1}^{\infty}P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(|\xi_n(\omega)-\xi(\omega)|\geqslant\frac{1}{m}\right)\right\}$$

可由(5.4.15)推出(5.4.14),这就说明了两种表达法的等价性.

利用概率的连续性可知,(5.4.15)等价于

$$\lim_{k\to\infty}P\left\{\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}=0 \qquad (5.4.16)$$

根据德莫根定理又知(5.4.15)等价于

$$\lim_{k\to\infty}P\left\{\bigcap_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|<\varepsilon)\right\}=1 \qquad (5.4.17)$$

由于

$$\{|\xi_k(\omega)-\xi(\omega)|\geqslant\varepsilon\}\subset\left\{\bigcup_{n=k}^{\infty}(|\xi_n(\omega)-\xi(\omega)|\geqslant\varepsilon)\right\}$$

因此若(5.4.16)成立,则

$$\lim_{k\to\infty}P\{|\xi_k(\omega)-\xi(\omega)|\geqslant\varepsilon\}=0$$

这样一来,我们已证得

定理 5.4.1 $\xi_n(\omega)\xrightarrow{a.s.}\xi(\omega)\Rightarrow\xi_n(\omega)\xrightarrow{P}\xi(\omega)$.

下例说明一般不能由依概率收敛推得以概率 1 收敛,所以以概率 1 收敛是比依概率收敛更强的一种收敛性.

[例 1] 取 $\Omega=(0,1]$,$\mathscr{F}$ 为(0,1]中博雷尔点集全体所构成的 σ 域,P 为勒贝格测度,令

$$\eta_{ki}(\omega)=\begin{cases}1, & \omega\in\left(\dfrac{i-1}{k},\dfrac{i}{k}\right] \quad i=1,2,\cdots,k\\ 0, & \omega\overline{\in}\left(\dfrac{i-1}{k},\dfrac{i}{k}\right] \quad k=1,2,\cdots\end{cases} \qquad (5.4.18)$$

定义

$\xi_1(\omega)=\eta_{11}(\omega),\xi_2(\omega)=\eta_{21}(\omega),\xi_3(\omega)=\eta_{22}(\omega)$

$\xi_4(\omega)=\eta_{31}(\omega),\xi_5(\omega)=\eta_{32}(\omega),\cdots$

一般 $\xi_n(\omega)=\eta_{ki}(\omega)$，其中 $n=i+\frac{k(k-1)}{2}$，这样定义的 $\{\xi_n(\omega)\}$ 是一列随机变量. 但对于任何一个 $\omega\in(0,1]$，$\xi_n(\omega))$ 必有无限个 k,i 使其取值 0，也有无限个 k,i 使其取值 1，因此 $\{\xi_n(\omega)\}$ 不是以概率 1 收敛于 0. 但是另一方面，对任意的 $\varepsilon>0$，

$$P\{|\eta_{ki}(\omega)|\geqslant\varepsilon\}\leqslant\frac{1}{k}$$

当 $n\to\infty$ 时，由 $n=\frac{k(k-1)}{2}+i\leqslant\frac{k(k-1)}{2}+k$，知道 $k\to\infty$，因此

$$\begin{aligned}&\lim_{n\to\infty}P\{|\xi_n(\omega)|\geqslant\varepsilon\}\\=&\lim_{n\to\infty}P\{|\eta_{ki}(\omega)|\geqslant\varepsilon\}=0\end{aligned}$$

所以 $\{\xi_n(\omega)\}$ 依概率收敛于 0.

不难验证，$\{\xi_n(\omega)\}$ 是 r 阶收敛于 0 的，因此例 1 也提供了 r 阶收敛推不出以概率 1 收敛之例.

我们以前讨论的大数定律只要求依概率收敛，若把收敛性要求提高为以概率 1 收敛，则得到的大数定律称为**强大数定律**(strong law of large numbers). 由定理 1 可知，若强大数定律成立，则通常的大数定律也一定成立，反之不然. 有时为区别起见，把依概率收敛意义下的大数定律称为**弱大数定律**(weak law of large numbers).

第一个强大数定律是由博雷尔在 1909 年对伯努利试验场合建立的.

二、博雷尔强大数定律

定理 5.4.2（博雷尔） 设 μ_n 是事件 A 在 n 次独立试验中的出现次数，在每次试验中事件 A 出现的概率均为 p，那么当 $n\to\infty$ 时，

$$P\left\{\frac{\mu_n}{n}\to p\right\}=1 \tag{5.4.19}$$

［证明］ 为使（5.4.19）成立，由（5.4.15）知，只须对任

意的 $\varepsilon>0$，成立

$$P\left\{\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}\left(\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right)\right\}=0 \tag{5.4.20}$$

若记 $A_n=\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}$，则上式可写成 $P\{\overline{\lim_{n\to\infty}}A_n\}=0$. 根据博雷尔—康特立引理，为证明（5.4.20）只要能证明级数

$$\sum_{n=1}^{\infty}P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\} \tag{5.4.21}$$

对任何 $\varepsilon>0$ 都收敛就可以了.

假如象证明伯努利大数定律那样用车贝晓夫不等式进行估计，只能得到

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{4n\varepsilon^2} \tag{5.4.22}$$

这对证明弱大数定律足够了，但为了保证（5.4.21）收敛还不行，这时必须寻找更好的估计式. 在这种特殊场合，马尔可夫不等式就够用了. 由于

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^4}E\left|\frac{\mu_n}{n}-p\right|^4 \tag{5.4.23}$$

问题是要计算$\frac{\mu_n}{n}$的四阶中心矩. 还是象过去一样，我们把 μ_n 表示成独立伯努利 0—1 变量之和，这样

$$\frac{\mu_n}{n}-p=\frac{1}{n}\sum_{i=1}^{n}(\xi_i-p)$$

所以

$$E\left(\frac{\mu_n}{n}-p\right)^4$$

$$=\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}E(\xi_i-p)(\xi_j-p)(\xi_k-p)(\xi_l-p)$$

注意到各 ξ_i 的独立性及 $E(\xi_i-p)=0$，因此上面的和式中只有 $E(\xi_i-p)^4$及 $E(\xi_i-p)^2(\xi_j-p)^2$ 的项才不等于 0，显然

$$E(\xi_i-p)^4=pq(p^3+q^3) \tag{5.4.24}$$

$$E(\xi_i-p)^2(\xi_j-p)^2=p^2q^2 \quad (i\neq j) \tag{5.4.25}$$

(5.4.24) 形式的项有 n 项，(5.4.25) 形式的项有 $\binom{4}{2}\binom{n}{2}=3n(n-1)$ 项，因此

$$E\left(\frac{\mu_n}{n}-p\right)^4=\frac{pq}{n^4}\left[n(p^3+q^3)+3pq(n^2-n)\right]<\frac{1}{4n^2} \tag{5.4.26}$$

于是

$$P\left\{\left|\frac{\mu_n}{n}-p\right|\geqslant\varepsilon\right\}<\frac{1}{4\varepsilon^4 n^2} \tag{5.4.27}$$

这个估计式已经比 (5.4.22) 进了一大步，它可以保证 (5.4.21) 收敛，从而证明了定理.

从本书第一节介绍随机事件频率稳定性时，我们就期待着这样一个结论，即当试验次数无限增加时，频率将趋于概率，博雷尔强大数定律正给出了这个结果. 从伯努利大数定律并不能引伸出这个结论，它只断言一个不等式 $\left|\frac{\mu_n}{n}-p\right|<\varepsilon$ 成立的概率可以大于 $1-\eta$，不论 η 是什么正数；但是事件

$$\left|\frac{\mu_{n+1}}{n+1}-p\right|\geqslant\varepsilon,\quad \left|\frac{\mu_{n+2}}{n+2}-p\right|\geqslant\varepsilon,\quad \cdots,\quad \left|\frac{\mu_{2n}}{2n}-p\right|\geqslant\varepsilon,\quad \cdots$$

中至少有一个发生仍是可能的，因为它是可列个事件之并，而我们只知道每个事件的概率很小. 但博雷尔强大数定律则断言 $\left|\frac{\mu_n}{n}-p\right|$ 以概率 1 变得很小，而且保持很小. 虽然从逻辑上讲，在投硬币时每次都出现正面是可能的，这时 $\frac{\mu_n}{n}=1$，因而 $\frac{\mu_n}{n}\to p$ 并不成立，但是强大数定律断言了这种事件发生的概率为 0.

三、科尔莫戈罗夫强大数定律

下面讨论更一般的强大数定律，先把其含义进一步明确如下：设 $\{\xi_i\}$ 是独立随机变量序列，若

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-E\xi_i)=0\right\}=1 \tag{5.4.28}$$

则称它满足**强大数定律**.

根据（5.4.16），这等价于要求对任意 $\varepsilon>0$ 成立

$$\lim_{m\to\infty}P\left\{\bigcup_{j=m}^{\infty}\left(\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right)\right\}=0 \qquad (5.4.29)$$

由于

$$\bigcup_{j=m}^{\infty}\left(\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right)$$

$$\subset\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\} \qquad (5.4.30)$$

因此需要对概率 $P\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$进行估计，这相当于在独立和场合对车贝晓夫不等式进行推广. 这方面已经有不少成果，在这里我们介绍一个由噶依克（Häjek）及瑞尼（Rēnyi）证明的不等式.

噶依克—瑞尼不等式 若 $\{\xi_i\}$ 是独立随机变量序列，$D\xi_i=\sigma_i^2<\infty$，$(i=1, 2, \cdots)$，而 $\{C_n\}$ 是一列正的非增常数序列，则对任意正整数 m，n $(m<n)$ 及 $\varepsilon>0$，均有

$$P\left\{\max_{m\leqslant j\leqslant n}C_j\left|\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left(C_m^2\sum_{j=1}^{m}\sigma_j^2+\sum_{j=m+1}^{n}C_j^2\sigma_j^2\right) \qquad (5.4.31)$$

［证明］ 记

$$S_k=\sum_{j=1}^{k}(\xi_j-E\xi_j)$$

及

$$\eta=\sum_{k=m}^{n-1}S_k^2(C_k^2-C_{k+1}^2)+C_n^2S_n^2 \qquad (5.4.32)$$

因此

$$\eta=\sum_{k=m}^{n}S_k^2C_k^2-\sum_{k=m}^{n-1}S_k^2C_{k+1}^2$$

$$= S_m^2 C_m^2 + \sum_{k=m+1}^{n} (S_k^2 - S_{k-1}^2) C_k^2$$

利用 $\{\xi_i\}$ 的独立性

$$ES_k^2 = \sum_{j=1}^{k} \sigma_j^2$$

所以有

$$E\eta = C_m^2 \sum_{j=1}^{m} \sigma_j^2 + \sum_{j=m+1}^{n} C_j^2 \sigma_j^2 \qquad (5.4.33)$$

对 $j=m$，$m+1$，…，n，记

$$E_j = \{C_k |S_k| < \varepsilon, \quad m \leqslant k < j; \quad C_j |S_j| \geqslant \varepsilon\} \qquad (5.4.34)$$

这样定义的 E_j（$j=m$，$m+1$，…，n）是互不相容的，而且

$$P\left\{ \max_{m\leqslant j\leqslant n} C_j \left| \sum_{i=1}^{j} (\xi_i - E\xi_i) \right| \geqslant \varepsilon \right\}$$

$$= P\left\{ \max_{m\leqslant j\leqslant n} C_j |S_j| \geqslant \varepsilon \right\} = \sum_{j=m}^{n} P(E_j) \qquad (5.4.35)$$

令

$$\chi_j(\omega) = \begin{cases} 1, & \omega \in E_j \\ 0, & \omega \overline{\in} E_j \end{cases} \quad j = m, m+1, \cdots, n \quad (5.4.36)$$

即 χ_j 是 E_j 的示性函数，注意到 E_j 的互不相容性及 $\sum_{j=m}^{n} E_j \subset \Omega$，因而

$$\sum_{j=m}^{n} \chi_j(\omega) \leqslant 1 \qquad (5.4.37)$$

所以有

$$E\eta \geqslant \sum_{j=m}^{n} E(\eta \chi_j) \qquad (5.4.38)$$

对 $j<k\leqslant n$，

$$S_k = S_j + (\xi_{j+1} - E\xi_{j+1}) + \cdots + (\xi_k - E\xi_k)$$

因此

$E(S_k^2 \chi_j) = E(S_j^2 \chi_j)$

$$+E\{[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]^2\cdot\chi_j\}$$
$$+2E\{S_j[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]\cdot\chi_j\}$$
$$\geqslant E(S_j^2\chi_j)+2E\{S_j\chi_j[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]\} \tag{5.4.39}$$

由于 $S_j\chi_j$ 只与 $\xi_1,\cdots,\xi_j$ 有关,因此与$[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]$独立,故

$$E\{S_j\chi_j[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]\}$$
$$=E(S_j\chi_j)\cdot E[(\xi_{j+1}-E\xi_{j+1})+\cdots+(\xi_k-E\xi_k)]=0 \tag{5.4.40}$$

而在 E_j 上 $|S_j|\geqslant\frac{\varepsilon}{C_j}$,故有

$$E(S_j^2\chi_j)\geqslant\frac{\varepsilon^2}{C_j^2}E\chi_j=\frac{\varepsilon^2}{C_j^2}P(E_j) \tag{5.4.41}$$

因此由(5.4.39),(5.4.40)及(5.4.41)得到

$$E(S_k^2\chi_j)\geqslant\frac{\varepsilon^2}{C_j^2}P(E_j),j\leqslant k\leqslant n \tag{5.4.42}$$

现在当 $m\leqslant j\leqslant n$ 时,由(5.4.32)知

$$E(\eta\chi_j)=\sum_{k=m}^{n-1}E(S_k^2\chi_j)(C_k^2-C_{k+1}^2)+C_n^2E(S_n^2\chi_j)$$
$$\geqslant\sum_{k=j}^{n-1}E(S_k^2\chi_j)(C_k^2-C_{k+1}^2)+C_n^2E(S_n^2\chi_j)$$
$$\geqslant\frac{\varepsilon^2}{C_j^2}P(E_j)\left[\sum_{k=j}^{n-1}(C_k^2-C_{k+1}^2)+C_n^2\right]$$
$$=\varepsilon^2P(E_j) \tag{5.4.43}$$

由(5.4.38)及(5.4.43)

$$E\eta\geqslant\varepsilon^2\sum_{j=m}^{n}P(E_j) \tag{5.4.44}$$

利用(5.4.33),(5.4.35)及(5.4.44)即得不等式(5.4.31).

在噶依克—瑞尼不等式中,特别令 $m=1,C_j=1$,则得到著名的科尔莫戈罗夫不等式.

科尔莫戈罗夫不等式 设 $\xi_1,\xi_2\cdots,\xi_n$ 是独立随机变量，方差有限，则对任意 $\varepsilon>0$，成立

$$P\left\{\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}\sum_{j=1}^{n}D\xi_j \quad (5.4.45)$$

科尔莫戈罗夫不等式是概率论中最重要的不等式之一，有广泛的应用. 在上式中，若令 $n=1$，则得到

$$P\{|\xi_1-E\xi_1|\geqslant\varepsilon\}\leqslant\frac{D\xi_1}{\varepsilon^2}$$

这正是车贝晓夫不等式. 因此科尔莫戈罗夫不等式是车贝晓夫不等式的推广，而噶依克—瑞尼不等式又是科尔莫戈罗夫不等式的推广.

利用噶依克—瑞尼不等式，能证明下面重要结果.

定理 5.4.3（科尔莫戈罗夫强大数定律） 设 $\{\xi_i\}$，$i=1,2,\cdots$ 是独立随机变量序列，且 $\sum_{n=1}^{\infty}\frac{D\xi_n}{n^2}<\infty$，则成立

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-E\xi_i)=0\right\}=1 \quad (5.4.46)$$

［证明］ 在噶依克—瑞尼不等式中，令 $C_j=\frac{1}{j}$，可以得到

$$P\left\{\max_{m\leqslant j\leqslant n}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{j=1}^{m}D\xi_j+\sum_{j=m+1}^{n}\frac{D\xi_j}{j^2}\right)$$

由概率的连续性

$$P\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$=\lim_{n\to\infty}P\left\{\max_{m\leqslant j\leqslant n}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}$$

$$\leqslant\frac{1}{\varepsilon^2}\left(\frac{1}{m^2}\sum_{j=1}^{m}D\xi_j+\sum_{j=m+1}^{\infty}\frac{D\xi_j}{j^2}\right) \quad (5.4.47)$$

因为$\sum_{j=1}^{\infty}\frac{D\xi_j}{j^2}<\infty$，故由(5.4.47)得到

$$\lim_{m\to\infty}P\left\{\sup_{j\geqslant m}\left|\frac{1}{j}\sum_{i=1}^{j}(\xi_i-E\xi_i)\right|\geqslant\varepsilon\right\}=0$$

这正等价于(5.4.46). 定理证毕.

科尔莫戈罗夫强大数定律的另外两种常见的证明方法见习题51～55.

显然，由科尔莫戈罗夫强大数定律很容易推出博雷尔强大数定律.

四、独立同分布场合的强大数定律

在这种特殊的场合，可以找到强大数定律成立的充要条件，这个结果也属于科尔莫戈罗夫.

定理 5.4.4(科尔莫戈罗夫) 设 $\xi_1,\xi_2,\cdots$是相互独立相同分布的随机变量序列，则

$$\frac{1}{n}(\xi_1+\xi_2+\cdots+\xi_n)\xrightarrow{a.s.}a \tag{5.4.48}$$

成立的充要条件是 $E\xi_i$ 存在且等于 a.

［证明］ 若 ξ 的分布函数为 $F(x)$，我们来证明不等式：

$$\sum_{n=1}^{\infty}P\{|\xi|\geqslant n\}\leqslant E|\xi|\leqslant 1+\sum_{n=1}^{\infty}P\{|\xi|\geqslant n\} \tag{5.4.49}$$

事实上，

$$E|\xi|=\int_{-\infty}^{\infty}|x|dF(x)=\sum_{k=0}^{\infty}\int_{k\leqslant|x|<k+1}|x|dF(x)$$

因此

$$\sum_{k=0}^{\infty}kP\{k\leqslant|\xi|<k+1\}\leqslant E|\xi|$$

$$\leqslant\sum_{k=0}^{\infty}(k+1)P\{k\leqslant|\xi|<k+1\}$$

现在有

$$\sum_{k=0}^{\infty} kP\{k \leqslant |\xi| < k+1\} = \sum_{n=1}^{\infty}\sum_{k=n}^{\infty} P\{k \leqslant |\xi| < k+1\}$$

$$= \sum_{n=1}^{\infty} P\{|\xi| \geqslant n\}$$

及

$$\sum_{k=0}^{\infty}(k+1)P\{k \leqslant |\xi| < k+1\}$$

$$= \sum_{k=0}^{\infty} kP\{k \leqslant |\xi| < k+1\} + 1 = \sum_{n=1}^{\infty} P\{|\xi| \geqslant n\} + 1$$

这就证得了(5.4.49). 这个不等式说明 $E|\xi|<\infty$的充要条件为

$$\sum_{n=1}^{\infty} P\{|\xi| \geqslant n\} < \infty$$

记 $S_n=\xi_1+\xi_2+\cdots+\xi_n$，若$\frac{S_n}{n}\xrightarrow{a.s.}\mu$，这里 μ 是有限数，则

$$\frac{\xi_n}{n} = \frac{S_n}{n} - \frac{n-1}{n}\cdot\frac{S_{n-1}}{n-1} \xrightarrow{a.s.} 0 \qquad (5.4.50)$$

这样一来，事件$\{|\xi_n|\geqslant n\}$发生无穷多次的概率为 0，因此注意到 ξ_i 的独立性，并利用博雷尔—康特立引理(ii)，可知

$$\sum_{n=1}^{\infty} P\{|\xi_n| \geqslant n\} < \infty$$

再由(5.4.49)即知 $E|\xi_i|<\infty$，这时显然有 $a=E\xi_i$，这样，我们已证得必要性.

下证充分性. 用“截尾法”，令

$$\xi_n^* = \begin{cases} \xi_n, & |\xi_n| < n \\ 0, & |\xi_n| \geqslant n \end{cases} \qquad (5.4.51)$$

先验证$\{\xi_n^*\}$满足科尔莫戈罗夫强大数定律条件. 以 $F(x)$记 ξ_i 的分布函数，则

$$D\xi_n^* \leqslant E\xi_n^{*2} = \int_{-n}^{n} x^2 \mathrm{d}F(x) \leqslant \sum_{k=1}^{n} k^2 P\{k-1 \leqslant |\xi_n| < k\}$$

$$\sum_{n=1}^{\infty}\frac{D\xi_n^*}{n^2}\leqslant\sum_{n=1}^{\infty}\sum_{k=1}^{n}\frac{k^2}{n^2}P\{k-1\leqslant|\xi_n|<k\}$$
$$=\sum_{k=1}^{\infty}\sum_{n=k}^{\infty}\frac{k^2}{n^2}P\{k-1\leqslant|\xi_n|<k\}$$
$$=\sum_{k=1}^{\infty}k^2P\{k-1\leqslant|\xi_n|<k\}\sum_{n=k}^{\infty}\frac{1}{n^2}$$

由于

$$\sum_{n=k}^{\infty}\frac{1}{n^2}<\frac{1}{k^2}+\sum_{n=k+1}^{\infty}\frac{1}{n(n-1)}=\frac{1}{k^2}+\frac{1}{k}\leqslant\frac{2}{k}$$

故

$$\sum_{n=1}^{\infty}\frac{D\xi_n^*}{n^2}<2\sum_{k=1}^{\infty}kP\{k-1\leqslant|\xi_n|<k\}<\infty$$

因此

$$P\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i^*-E\xi_i^*)=0\}=1 \qquad (5.4.52)$$

因为

$$E\xi_n^*=\int_{-n}^{n}x\mathrm{d}F(x)$$

显然$\lim\limits_{n\to\infty}E\xi_n^*=E\xi_1=a$，因此$\lim\limits_{n\to\infty}\frac{1}{n}\sum\limits_{i=1}^{n}E\xi_i^*=E\xi_1=a$，由于

$$\left|\frac{1}{n}\sum_{i=1}^{n}(\xi_i-a)\right|\leqslant\left|\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\xi_i^*)\right|$$
$$+\left|\frac{1}{n}\sum_{i=1}^{n}(\xi_i^*-E\xi_i^*)\right|+\left|\frac{1}{n}\sum_{i=1}^{n}(E\xi_i^*-a)\right| \qquad (5.4.53)$$

为证（5.4.48）成立，只须再证$\frac{1}{n}\sum\limits_{i=1}^{n}(\xi_i-\xi_i^*)\xrightarrow{a.\ s.}0$. 然而

$$\sum_{i=1}^{\infty}P\{\xi_i\neq\xi_i^*\}=\sum_{i=1}^{\infty}P\{|\xi_i|\geqslant i\}\leqslant E|\xi_1|<\infty$$

由博雷尔—康特立引理知，以概率 1 有

$\xi_i(\omega)\neq\xi_i^*(\omega)$,只对有限个 i 成立

因此

$$P\left\{\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}(\xi_i-\xi_i^*)=0\right\}=1$$

这样，定理的证明已经完成.

显然，科尔莫戈罗夫的这个结果是辛钦大数定律的加强，只有它才能保证在每次试验中当 $n\to\infty$时，子样的均值$\frac{1}{n}\sum_{i=1}^{n}\xi_i$将最终地趋于母体的均值；当然从逻辑上讲也有可能失败，但是这种不愉快场合发生的概率等于 0. 用蒙特卡洛方法计算积分所需要的正是强大数定律.

* §5. 中心极限定理

一、林德贝格条件与费勒（Feller）条件

本节将最后解决古典的中心极限定理. 为此先把问题的提法作进一步明确.

古典的中心极限定理讨论的是独立和的分布函数向正态分布收敛的最普遍条件. 这个问题一方面可以看作是棣莫弗—拉普拉斯古典结果的一般化，另一方面也解释了正态分布为什么是最常见的一种分布.

自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见. 例如炮弹的弹落点服从正态分布，人的许多生理特征如身长、体重等也服从正态分布. 观察表明，如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不很大，则这种量通常都服从或近似服从正态分布.

另外，在数理统计中，经常都假定母体服从正态分布，这也要求通过对中心极限定理的研究来阐明这个假定的正确性和适用条件.

现在,这个问题从某种意义上来讲已经得到了最后解决. 1922

年林德贝格提出了充分条件；1935年，费勒进一步指出，在某种条件下，这个条件也是必要的．这样就搞清了向正态分布收敛的充要条件．下面就介绍这些条件．

设 $\xi_1, \xi_2\cdots, \xi_n, \cdots$ 是一个相互独立的随机变量序列，它们具有有限的数学期望和方差：

$$a_k = E\xi_k, \quad b_k^2 = D\xi_k \quad (k = 1,2,\cdots,n,\cdots)$$

记

$$B_n^2 = \sum_{k=1}^{n} b_k^2$$

作和数

$$\zeta_n = \sum_{k=1}^{n} \frac{\xi_k - a_k}{B_n} \tag{5.5.1}$$

我们需要寻找和数 ζ_n 的分布函数趋于正态分布函数的充要条件．

与独立同分布场合比较，这里保留了独立性的假定，但是去掉了同分布的要求．今后我们将以 $F_k(x)$ 记 ξ_k 的分布函数．显然为了讨论 ζ_n 的极限分布，要使问题的提法有意义，对各个加项必须有一定要求．例如若允许从第二项起都等于0，则极限分布显然由 $F_1(x)$ 完全确定，这时就很难有什么有意思的结果．排除这个困难的办法是规定加项中不能有某些项起支配作用，在实际工作中人们就是这样处理的，例如为了讨论测量的随机误差，总预先把一些系统性的误差先扣除掉．

为了使极限分布是正态分布，还要求各个加项"均匀地小"，怎样明确表达这个要求呢？下面先作一个启发性的推导．

设 A_k 表示下述事件：

$$\{|\xi_k - a_k| > \tau B_n\} \quad (k = 1,2,\cdots,n)$$

则有

$$P\{\max_{1\leqslant k\leqslant n} |\xi_k - a_k| > \tau B_n\} = P\left\{\bigcup_{k=1}^{n}(|\xi_k - a_k| > \tau B_n)\right\}$$

$$= P\{A_1 \cup A_2 \cup \cdots \cup A_n\} \leqslant \sum_{k=1}^{n} P(A_k)$$

$$= \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} \mathrm{d}F_k(x)$$

$$\leqslant \frac{1}{(\tau B_n)^2} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 \mathrm{d}F_k(x)$$

$$= \frac{1}{\tau^2 B_n^2} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 \mathrm{d}F_k(x)$$

因此，只要对于任何 $\tau>0$，成立

$$\lim_{n\to\infty} \frac{1}{B_n^2} \sum_{k=1}^{n} \int_{|x-a_k|>\tau B_n} (x-a_k)^2 \mathrm{d}F_k(x) = 0 \qquad (5.5.2)$$

就可以保证总和(5.5.1)中各加项“均匀地小”. 上述条件(5.5.2)称为**林德贝格条件**. 林德贝格证明了条件(5.5.2)是和数(5.5.1)的分布函数趋于正态分布函数的充分条件.

但是林德贝格条件不是中心极限定理成立的必要条件（参看习题 58）. 不过，费勒进一步指出，假如下面条件得到满足：

$$\lim_{n\to\infty} \max_{k\leqslant n} \frac{b_k}{B_n} = 0 \qquad (5.5.3)$$

则林德贝格条件也是中心极限定理成立的必要条件.

条件（5.5.3）称为**费勒条件**. 下面考察一下费勒条件的含义.

定理 5.5.1 费勒条件（5.5.3）等价于

$$\lim_{n\to\infty} B_n = \infty \qquad (5.5.4)$$

$$\lim_{n\to\infty} \frac{b_n}{B_n} = 0 \qquad (5.5.5)$$

［证明］ 若（5.5.3）成立，则由 $\frac{b_n}{B_n} \leqslant \max_{k\leqslant n} \frac{b_k}{B_n}$ 立刻得到（5.5.5）；又若 $B_n \to B$，（$B<\infty$），不妨假定 $b_1>0$，则因 $\max_{k\leqslant n} \frac{b_k}{B_n} \geqslant \frac{b_1}{B_n}$，故 $\lim_{n\to\infty} \max_{k\leqslant n} \frac{b_k}{B_n} \geqslant \frac{b_1}{B} > 0$，这与（5.5.3）矛盾，因此应有（5.5.4）.

反之，设（5.5.4）（5.5.5）成立. 对任意 $\varepsilon>0$，存在正整数

M，使$\frac{b_k}{B_k}<\varepsilon$对一切$k>M$成立．固定$M$之后，由于（5.5.4），可以选一个正整数$N\geqslant M$，使$\max\limits_{k\leqslant M}\frac{b_k}{B_N}<\varepsilon$，下证对一切$n\geqslant N$均有

$$\max_{k\leqslant n}\frac{b_k}{B_n}<\varepsilon \tag{5.5.6}$$

事实上，利用B_n的单调不减性，对一切$n\geqslant N\geqslant M$有

$$\max_{k\leqslant M}\frac{b_k}{B_n}\leqslant\max_{k\leqslant M}\frac{b_k}{B_N}<\varepsilon$$

$$\max_{M<k\leqslant n}\frac{b_k}{B_n}\leqslant\max_{M<k\leqslant n}\frac{b_k}{B_k}<\varepsilon$$

因此（5.5.6）成立，这就证得了（5.5.3），定理证毕．

量$\frac{b_k}{B_n}$可以看作是分量ξ_k对总和ζ_n的贡献，因此费勒条件相当于说：总和是大量“可忽略的”分量之和．

下面我们转入证明主要定理．

二、林德贝格—费勒定理

为了不打断主要定理的证明，我们把在定理证明中要用到的若干事实，以引理的形式给出．

引理 5.5.1 对$n=1,2,\cdots$及任意的t，

$$\left|\mathrm{e}^{\mathrm{i}t}-1-\frac{\mathrm{i}t}{1!}-\cdots-\frac{(\mathrm{i}t)^{n-1}}{(n-1)!}\right|\leqslant\frac{|t|^n}{n!} \tag{5.5.7}$$

［证明］记

$$g_n(t)=\mathrm{e}^{\mathrm{i}t}-1-\frac{\mathrm{i}t}{1!}-\cdots-\frac{(\mathrm{i}t)^{n-1}}{(n-1)!}$$

先设$t>0$，由于

$$g_1(t)=\mathrm{i}\int_0^t\mathrm{e}^{\mathrm{i}x}\mathrm{d}x \tag{5.5.8}$$

因此$|g_1(t)|\leqslant t$，其次，对$n>1$

$$g_n(t)=\mathrm{i}\int_0^t g_{n-1}(x)\mathrm{d}x \tag{5.5.9}$$

用归纳法即得（5.5.7）.

由于$|\overline{g_n(t)}|=|g_n(t)|$，因此（5.5.7）对$t<0$也成立. $t=0$结论显然成立.

特别地，我们要用到

$$|e^{it}-1-it|\leqslant\frac{|t|^2}{2} \tag{5.5.10}$$

$$\left|e^{it}-1-it+\frac{t^2}{2}\right|\leqslant\frac{|t|^3}{6} \tag{5.5.11}$$

类似地可以得到，对$t>0$

$$1-\cos t=\left|\int_0^t\sin x\mathrm{d}x\right|\leqslant\int_0^t|\sin x|\mathrm{d}x\leqslant\int_0^t x\mathrm{d}x=\frac{t^2}{2} \tag{5.5.12}$$

这式两边都是t的偶函数，故显然对$t\leqslant 0$也成立.

引理 5.5.2 对于任何满足$|a_k|\leqslant 1$及$|b_k|\leqslant 1$ $(k=1,2,\cdots,n)$的复数，有

$$|a_1a_2\cdots a_n-b_1b_2\cdots b_n|\leqslant\sum_{k=1}^{n}|a_k-b_k| \tag{5.5.13}$$

［证明］ 显然

$$a_1a_2-b_1b_2=(a_1-b_1)a_2+(a_2-b_2)b_1$$

因此

$$|a_1a_2-b_1b_2|\leqslant|a_1-b_1|+|a_2-b_2|$$

用归纳法即得（5.5.13）

引理 5.5.3 若$\varphi(t)$是特征函数，则$e^{\varphi(t)-1}$也是特征函数，特别地

$$|e^{\varphi(t)-1}|\leqslant 1 \tag{5.5.14}$$

［证明］ 定义随机变量

$$\eta=\xi_1+\xi_2+\cdots+\xi_\nu$$

其中ξ_1，ξ_2，…相互独立，均有特征函数$\varphi(t)$，ν服从参数$\lambda=1$的泊松分布，且与诸ξ_i独立，不难验证η的特征函数为$e^{\varphi(t)-1}$，由特征函数的性质即知（5.5.14）成立.

现在叙述并证明主要的结果.

定理 5.5.2 对（5.5.1）中定义的和数ζ_n，成立

$$\lim_{n\to\infty}P\{\zeta_n<x\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t \qquad (5.5.15)$$

与费勒条件（5.5.3）的充要条件是林德贝格条件（5.5.2）成立.

［证明］ 为书写方便起见，我们引用记号

$$\xi_{nk}=\frac{\xi_k-a_k}{B_n} \qquad (5.5.16)$$

显然

$$E\xi_{nk}=0,\qquad D\xi_{nk}=\frac{D\xi_k}{B_n^2}=\frac{b_k^2}{B_n^2} \qquad (5.5.17)$$

$$\sum_{k=1}^{n}D\xi_{nk}=\frac{1}{B_n^2}\sum_{k=1}^{n}D\xi_k=1 \qquad (5.5.18)$$

以 $f_{nk}(t)$及 $F_{nk}(x)$分别表示 ξ_{nk}的特征函数与分布函数，那么

$$F_{nk}(x)=P\left\{\frac{\xi_k-a_k}{B_n}<x\right\}=F_k(B_nx+a_k) \qquad (5.5.19)$$

这时

$$\frac{1}{B_n^2}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\mathrm{d}F_k(x)$$

$$=\int_{\left|\frac{x-a_k}{B_n}\right|>\tau}\left(\frac{x-a_k}{B_n}\right)^2\mathrm{d}F_k(x)$$

$$=\int_{|y|>\tau}y^2\mathrm{d}F_{nk}(y)$$

因此林德贝格条件（5.5.2）化为：对任意 $\tau>0$,

$$\lim_{n\to\infty}\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x)=0 \qquad (5.5.20)$$

现在开始证明定理. 设 t 是任意固定的实数.

为证（5.5.15）必须证明

$$f_{n1}(t)\cdots f_{nn}(t)\to \mathrm{e}^{-t^2/2} \qquad (5.5.21)$$

我们先证明，在费勒条件（5.5.3）成立的假定下，（5.5.21）与

下式是等价的：

$$\sum_{k=1}^{n}(f_{nk}(t)-1)+\frac{1}{2}t^2\to 0 \tag{5.5.22}$$

事实上，若（5.5.3）成立，是对任意 $\varepsilon>0$，只要 n 充分大，均有

$$\frac{b_k}{B_n}<\varepsilon,\quad k=1,2,\cdots,n$$

另一方面，由（5.5.10）可知存在复数 θ，使得

$$\mathrm{e}^{\mathrm{i}tx}-1-\mathrm{i}tx=\theta\frac{(tx)^2}{2},\quad |\theta|\leqslant 1 \tag{5.5.23}$$

因此

$$f_{nk}(t)-1-\mathrm{i}tE\xi_{nk}=\frac{\theta t^2}{2}\int_{-\infty}^{\infty}x^2\mathrm{d}F_{nk}(x)$$

再由（5.5.17）可得：

$$\begin{aligned}|f_{nk}(t)-1|&=\left|\frac{\theta t^2}{2}\int_{-\infty}^{\infty}x^2\mathrm{d}F_{nk}(x)\right|\\&\leqslant\frac{t^2}{2}\int_{-\infty}^{\infty}x^2\mathrm{d}F_{nk}(x)\\&=\frac{t^2}{2}\frac{b_k^2}{B_n^2}<\frac{1}{2}\varepsilon^2t^2\end{aligned} \tag{5.5.24}$$

对任意 $\delta>0$，只要 $|z|$ 充分小，就可以有

$$|\mathrm{e}^z-1-z|<\delta|z| \tag{5.5.25}$$

因此由引理 5.5.3、引理 5.5.2 及（5.5.24）、（5.5.25），只要 n 充分大，就有

$$\begin{aligned}&\left|\mathrm{e}^{\sum_{k=1}^{n}[f_{nk}(t)-1]}-f_{n1}(t)\cdots f_{nn}(t)\right|\\&\leqslant\sum_{k=1}^{n}\left|\mathrm{e}^{f_{nk}(t)-1}-f_{nk}(t)\right|\\&\leqslant\delta\sum_{k=1}^{n}|f_{nk}(t)-1|\\&\leqslant\frac{1}{2}\delta t^2\sum_{k=1}^{n}\frac{b_k^2}{B_n^2}=\frac{1}{2}\delta t^2\end{aligned} \tag{5.5.26}$$

因为 δ 可以任意小，故左边趋于 0，因此证得（5.5.21）与（5.5.22）的等价性.

接着证明定理的充分性. 先证由林德贝格条件可以推出费勒条件. 事实上，

$$
\begin{aligned}
\frac{b_k^2}{B_n^2} &= \int_{-\infty}^{\infty} x^2 \mathrm{d}F_{nk}(x) \\
&= \int_{|x|\leqslant\tau} x^2 \mathrm{d}F_{nk}(x) + \int_{|x|>\tau} x^2 \mathrm{d}F_{nk}(x) \\
&\leqslant \tau^2 + \sum_{k=1}^{n} \int_{|x|>\tau} x^2 \mathrm{d}F_{nk}(x) \qquad (5.5.27)
\end{aligned}
$$

右边与 k 无关，而且 τ 可选得任意地小；对选定的 τ，由林德贝格条件（5.5.20）知道第二式当 n 足够大时也可任意地小. 这样，费勒条件成立.

其次证明林德贝格条件能保证(5.5.15)成立. 注意到(5.5.17)及(5.5.18)，可知

$$
\begin{aligned}
&\sum_{k=1}^{n} [f_{nk}(t) - 1] + \frac{1}{2}t^2 \\
&= \sum_{k=1}^{n} \int_{-\infty}^{\infty} \left[\mathrm{e}^{\mathrm{i}tx} - 1 - \mathrm{i}tx + \frac{t^2x^2}{2} \right] \mathrm{d}F_{nk}(x)
\end{aligned}
$$

利用（5.5.11），当 $|x|\leqslant\tau$ 时，

$$
\left| \mathrm{e}^{\mathrm{i}tx} - 1 - \mathrm{i}tx + \frac{t^2x^2}{2} \right| \leqslant \frac{|tx|^3}{6} \leqslant \frac{\tau |t|^3 x^2}{6}
$$

又利用(5.5.10)，当 $|x|>\tau$ 时，

$$
\left| \mathrm{e}^{\mathrm{i}tx} - 1 - \mathrm{i}tx + \frac{t^2x^2}{2} \right| \leqslant |\mathrm{e}^{\mathrm{i}tx} - 1 - \mathrm{i}tx| + \frac{t^2x^2}{2} \leqslant t^2x^2
$$

因此

$$
\begin{aligned}
&\left| \sum_{k=1}^{n} [f_{nk}(t) - 1] + \frac{1}{2}t^2 \right| \\
&\leqslant \sum_{k=1}^{n} \int_{|x|\leqslant\tau} \frac{\tau |t|^3 x^2}{6} \mathrm{d}F_{nk}(x) + \sum_{k=1}^{n} \int_{|x|>\tau} t^2x^2 \mathrm{d}F_{nk}(x)
\end{aligned}
$$

$$\leqslant \frac{\tau|t|^3}{6}\sum_{k=1}^{n}\int_{-\infty}^{\infty}x^2\mathrm{d}F_{nk}(x)+t^2\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x)$$

$$=\frac{\tau|t|^3}{6}+t^2\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x) \tag{5.5.28}$$

对任给 $\varepsilon>0$，由于 τ 的任意性，可选得使 $\frac{\tau|t|^3}{6}<\frac{\varepsilon}{2}$，对选定的 τ，用林德贝格条件知只要 n 充分大，也可使 $t^2\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x)<\frac{\varepsilon}{2}$，因此我们已证得了（5.5.22)，但由于我们已证过费勒条件（5.5.3）成立，这时（5.5.22）与（5.5.21）是等价的，因而（5.5.21）也成立，根据特征函数连续性定理可知（5.5.15）成立.

再证定理的必要性.

由于(5.5.15)成立,因此相应的特征函数应满足(5.5.21).但在费勒条件成立时,这又推出(5.5.22),因此

$$\sum_{k=1}^{n}[f_{nk}(t)-1]+\frac{t^2}{2}$$
$$=\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[\mathrm{e}^{itx}-1+\frac{t^2x^2}{2}\right]\mathrm{d}F_{nk}(x)\to 0 \tag{5.5.29}$$

因为由(5.5.12)可得 $\cos tx-1+\frac{t^2x^2}{2}\geqslant 0$,因此上述被积函数的实部是非负的,故

$$\mathrm{Re}\left(\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[\mathrm{e}^{itx}-1+\frac{t^2x^2}{2}\right]\mathrm{d}F_{nk}(x)\right)$$
$$=\sum_{k=1}^{n}\int_{-\infty}^{\infty}\left[\cos tx-1+\frac{t^2x^2}{2}\right]\mathrm{d}F_{nk}(x)$$
$$\geqslant\sum_{k=1}^{n}\int_{|x|>\tau}\left[\cos tx-1+\frac{t^2x^2}{2}\right]\mathrm{d}F_{nk}(x)$$
$$\geqslant\sum_{k=1}^{n}\int_{|x|>\tau}\left(\frac{t^2x^2}{2}-2\right)\mathrm{d}F_{nk}(x)$$
$$=\frac{t^2}{2}\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x)-2\sum_{k=1}^{n}\int_{|x|>\tau}\mathrm{d}F_{nk}(x)$$

$$\geqslant\left(\frac{t^2}{2}-2\tau^{-2}\right)\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x) \tag{5.5.30}$$

因为对任意 $\tau>0$，可找到 t，使 $\frac{t^2}{2}-2\tau^{-2}>0$，这时由（5.5.29），（5.5.30）可得

$$\left(\frac{t^2}{2}-2\tau^{-2}\right)\sum_{k=1}^{n}\int_{|x|>\tau}x^2\mathrm{d}F_{nk}(x)\to 0$$

故林德贝格条件成立，定理证毕.

三、若干推论

林德贝格条件给出了中心极限定理成立的普遍条件，由它可以推出许多特殊的结果.

首先，我们来说明独立同分布场合的林德贝格—莱维定理是定理 5.5.2 的特例.

若 ξ_1，ξ_2，…是独立同分布随机变量序列，$E\xi_k=a$，$0<\sigma^2=D\xi_k<\infty$，则

$$B_n=\sqrt{n}\,\sigma \tag{5.5.31}$$

这时

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\mathrm{d}F_k(x)$$

$$=\frac{1}{n\sigma^2}\cdot n\int_{|x-a|>\tau\sigma\sqrt{n}}(x-a)^2\mathrm{d}F(x) \tag{5.5.32}$$

由于方差 $0<\sigma^2<\infty$，上式右边的积分当 $n\to\infty$ 时趋于 0，故林德贝格条件得到满足，所以中心极限定理成立.

下面我们再来给出两个有用的结果.

定理 5.5.3 若 $\xi_1,\xi_2,\cdots$是独立随机变量序列，存在常数 K_n，使 $\max\limits_{1\leqslant j\leqslant n}|\xi_j|\leqslant K_n$（$n=1,2,\cdots$），且 $\lim\limits_{n\to\infty}\frac{K_n}{B_n}=0$，则

$$P\left\{\sum_{k=1}^{n}\frac{\xi_k-a_k}{B_n}<x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t$$

［证明］ 由假定，对任意的 $\varepsilon>0$，只要 n 充分大就有 $2K_n\leqslant\varepsilon B_n$，显然

$$|\xi_j-a_j|\leqslant 2K_n,\quad j=1,2,\cdots,n$$

因此

$$\{|\xi_j-a_j|\leqslant\varepsilon B_n\}=\Omega\quad(1\leqslant j\leqslant n)\tag{5.5.33}$$

所以

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{j=1}^{n}\int_{|x-a_j|\leqslant\varepsilon B_n}(x-a_j)^2\mathrm{d}F_j(x)$$

$$=\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{j=1}^{n}\int_{-\infty}^{\infty}(x-a_j)^2\mathrm{d}F_j(x)=1$$

因此林德贝格条件得到满足，所以中心极限定理成立.

定理 5.5.4（李雅普诺夫） 如果对相互独立的随机变量序列 $\xi_1,\xi_2,\cdots,\xi_n,\cdots$ 能选择这样一个正数 $\delta>0$，使当 $n\to\infty$ 时，

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E|\xi_k-a_k|^{2+\delta}\to 0\tag{5.5.34}$$

则

$$P\left\{\frac{1}{B_n}\sum_{k=1}^{n}(\xi_k-a_k)<x\right\}\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-t^2/2}\mathrm{d}t$$

［证明］ 只要验证林德贝格条件就行了. 事实上，

$$\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2\mathrm{d}F_k(x)$$

$$\leqslant\frac{1}{B_n^2(\tau B_n)^\delta}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}|x-a_k|^{2+\delta}\mathrm{d}F_k(x)$$

$$\leqslant\frac{1}{\tau^\delta}\cdot\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}\int_{-\infty}^{\infty}|x-a_k|^{2+\delta}\mathrm{d}F_k(x)\to 0\quad(n\to\infty)$$

用定理 5.5.2 就可推得所需的结论.

第五章小结

本章研究了极限定理，这是概率论基础中比较深入的结果；前

几章学到的知识在这里得到了综合应用；一些重要问题在这里进一步讨论并获得解决.

在我们的课程中，为了使读者对极限定理有直观的认识，是从伯努利试验场合开始叙述的．这里所用的工具比较初等：伯努利大数定律是用矩法证明的；棣莫弗—拉普拉斯定理则通过利用斯特灵公式进行渐近估计而得到.

接着我们处理独立同分布场合，这是伯努利试验的直接推广，也是在实际中，特别是数理统计中，最常碰到的情况．为了证明辛钦大数定律及林德贝格—莱维定理已用到特征函数，所用的方法具有普遍性及简明性，是读者比较容易理解的．应当指出，收敛性概念及特征函数这一工具是深入研究极限定理所不可缺少的，所以对这部分内容我们作了相当详细的叙述．后面用到的主要是结论，因此有关证明在初学时不妨略去.

最后，介绍了强大数定律及一般场合的中心极限定理，这是概率论中相当深刻的结果．前者的证明通过建立比车贝晓夫不等式更为锐利的不等式而实现；后者的证明则得力于特征函数这一有力的工具的巧妙应用．到此为止，概率论中提出的古典极限定理问题已获得了令人满意的解决.

纵览整个发展过程，令人印象深刻的是，经典问题的最终解决，主要靠工具的改进，从直接展开，到矩法，再到特征函数法；也靠方法的精密化，例如从车贝晓夫不等式，到马尔可夫不等式，再到科尔莫戈罗夫等更精细的不等式.

习 题 五

1. ξ 为非负随机变量，若 $Ee^{a\xi}<\infty,(a>0)$，则对任意 $x>0, P\{\xi\geqslant x\}\leqslant e^{-ax}Ee^{a\xi}$.

2. 若 $h(x)\geqslant 0$，ξ 为随机变量，且 $Eh(\xi)<\infty$，则关于任何 $C>0, P\{h(\xi)\geqslant C\}\leqslant C^{-1}Eh(\xi)$.

*3.（单边车贝晓夫不等式） 设 ξ 为随机变量，$E\xi=0$，$D\xi=\sigma^2<\infty$，则

对任何一个 $a>0$，试证

$$P\{\xi>a\}\leqslant\frac{\sigma^2}{\sigma^2+a^2}$$

4. $\{\xi_k\}$ 各以$\frac{1}{2}$概率取值 k^s 和 $-k^s$，试证当 $s<\frac{1}{2}$时，大数定律可用于独立随机变量序列 ξ_1，…，ξ_n，…的算术平均值.

5. 若 ξ_k 的分布列为$\begin{pmatrix}\sqrt{\ln k} & -\sqrt{\ln k}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$，试证大数定律适用于独立随机变量序列 $\{\xi_k\}$.

6. 验证概率分布如下给定的独立随机变量序列是否满足马尔可夫条件：

(1) $P\{X_k=\pm 2^k\}=\frac{1}{2}$；

(2) $P\{X_k=\pm 2^k\}=2^{-(2k+1)},P\{X_k=0\}=1-2^{-2k}$；

(3) $P\{X_k=\pm k\}=\frac{1}{2}k^{-\frac{1}{2}},P\{X_k=0\}=1-k^{-\frac{1}{2}}$.

7. 若 ξ_k 具有有限方差，服从同一分布，但各 k 间，ξ_k 和 ξ_{k+1}有相关，而 ξ_k，ξ_l（$|k-l|\geqslant 2$）是独立的，证明这时对 $\{\xi_k\}$ 大数定律成立.

8.（伯恩斯坦定理）已知随机变量序列 ξ_1，ξ_2，…的方差有界：$D\xi_n\leqslant C$，并且当$|i-j|\to\infty$时，相关系数 $r_{ij}\to 0$，证明对 $\{\xi_n\}$ 成立大数定律.

*9.（格涅坚科定理）对随机变量序列 $\{\xi_i\}$，若记 $\eta_n=\frac{1}{n}(\xi_1+\cdots+\xi_n)$，$a_n=\frac{1}{n}(E\xi_1+\cdots+E\xi_n)$，则 $\{\xi_i\}$ 服从大数定律的充要条件是

$$\lim_{n\to\infty}E\left\{\frac{(\eta_n-a_n)^2}{1+(\eta_n-a_n)^2}\right\}=0$$

10. 用斯特林公式证明：当 $n\to\infty$，$m\to\infty$，$n-m\to\infty$，而$\frac{m}{n}\to 0$时，

$$\binom{2n}{n-m}\left(\frac{1}{2}\right)^{2n}\sim\frac{1}{\sqrt{\pi n}}\mathrm{e}^{-m^2/n}$$

11. 用（5.1.27）计算 b（5；500，0.01）及 b（40；10000，0.005）并与精确值比较.

12. 某计算机系统有 120 个终端，每个终端有 5%时间在使用，若各个终端使用与否是相互独立的，试求有 10 个或更多终端在使用的概率.

*13. 求证，在 $x>0$ 时，有不等式

$$\frac{x}{1+x^2}e^{-x^2/2} \leqslant \int_x^{\infty} e^{-t^2/2}dt \leqslant \frac{1}{x}e^{-x^2/2}$$

14. 用棣莫弗一拉普拉斯定理证明，在伯努利试验中，若 $0<p<1$，则不管 K 是如何大的常数，总有

$$P\{|\mu_n - np| < K\} \to 0 \quad (n\to\infty)$$

15. 用车贝晓夫不等式确定当掷一均匀铜币时，需投多少次才能保证使得正面出现的频率在 0.4 至 0.6 之间的概率不小于 90%，并用正态逼近计算同一问题.

16. 用车贝晓夫不等式及棣莫弗一拉普拉斯极限定理估计下面概率：

$$P\left\{\left|\frac{\mu_n}{n} - p\right| \geqslant \epsilon\right\}$$

并进行比较. 这里 μ_n 是 n 次伯努利试验中成功总次数，p 为每次成功的概率.

17. 现有一大批种子，其中良种占 1/6，今在其中任选 6000 粒，试问在这些种子中良种所占的比例与 1/6 之差小于 1%的概率是多少？

18. 种子中良种占 1/6，我们有 99%的把握断定在 6000 粒种子中良种所占的比例与 1/6 之差是多少？这时相应的良种粒数落在哪个范围内.

19. 蒲丰试验中掷铜币 4040 次，出正面 2048 次，试计算当重复蒲丰试验时，正面出现的频率与概率之差的偏离程度不大于蒲丰试验中所发生的偏差的概率.

*20. 设分布函数列 $\{F_n(x)\}$ 弱收敛于连续的分布函数 $F(x)$，试证这收敛对 $x\in \mathbf{R}^1$ 是一致的.

*21. 设 $\{F_n(x)\}$ 为一列正态分布函数，收敛于分布函数 $F(x)$，试证 $F(x)$ 也是正态分布函数.

*22. 试证若正态随机变量序列依概率收敛，则其数学期望与方差也收敛.

*23. 若 X_n 为多维正态随机向量，$X_n \xrightarrow{P} X$，试证 X 为正态向量.

24. 若 X_n 的概率分布为 $\begin{pmatrix} 0 & n \\ 1-\frac{1}{n} & \frac{1}{n} \end{pmatrix}$，试证相应的分布函数列收敛，但矩不收敛.

*25. 随机变量序列 $\{\xi_n\}$ 具有分布函数列 $\{F_n(x)\}$，且 $F_n(x) \to F(x)$，又 $\{\eta_n\}$ 依概率收敛于常数 $C>0$，试证：(1) $\zeta_n = \xi_n + \eta_n$ 的分布函数收敛于 $F(x-C)$；(2) $\zeta_n = \frac{\xi_n}{\eta_n}$ 的分布函数收敛于 $F(Cx)$.

*26. 试证：(1) $X_n \xrightarrow{P} X \Rightarrow X_n - X \xrightarrow{P} 0$；

(2) $X_n \xrightarrow{P} X$，$X_n \xrightarrow{P} Y \Rightarrow P\{X=Y\}=1$；

(3) $X_n \xrightarrow{P} X \Rightarrow X_n - X_m \xrightarrow{P} 0$ $(n, m \to \infty)$；

(4) $X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y \Rightarrow X_n \pm Y_n \xrightarrow{P} X \pm Y$；

(5) $X_n \xrightarrow{P} X$，k 是常数 $\Rightarrow kX_n \xrightarrow{P} kX$；

(6) $X_n \xrightarrow{P} X \Rightarrow X_n^2 \xrightarrow{P} X^2$；

(7) $X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，a，b 是常数 $\Rightarrow X_n Y_n \xrightarrow{P} ab$；

(8) $X_n \xrightarrow{P} 1 \Rightarrow X_n^{-1} \xrightarrow{P} 1$；

(9) $X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，a，b 是常数，$b \neq 0 \Rightarrow X_n Y_n^{-1} \xrightarrow{P} ab^{-1}$；

(10) $X_n \xrightarrow{P} X$，Y 是随机变量 $\Rightarrow X_n Y \xrightarrow{P} XY$；

(11) $X_n \xrightarrow{P} X$，$Y_n \xrightarrow{P} Y \Rightarrow X_n Y_n \xrightarrow{P} XY$.

27. 设 $X_n \xrightarrow{P} X$，而 g 是 $\mathbf{R}^1$ 上的连续函数，试证 $g(X_n) \xrightarrow{P} g(X)$.

28. 若 $\{X_n\}$ 是单调下降的正随机变量序列，且 $X_n \xrightarrow{P} 0$，试证 $X_n \xrightarrow{a.s.} 0$.

29. 若 $X_1, X_2, \cdots$ 是独立随机变量序列，μ 是整值随机变量，$P\{\mu=k\}=p_k$，且与 $\{X_i\}$ 独立，求 $\eta = X_1 + X_2 + \cdots + X_\mu$ 的特征函数.

30. 若 $f(t)$ 是非负定函数，试证：(1) $f(0)$ 是实的，且 $f(0) \geqslant 0$；(2) $f(-t) = \overline{f(t)}$；(3) $|f(t)| \leqslant f(0)$.

31. 用特征函数法直接证明棣莫弗--拉普拉斯积分极限定理.

32. 若飞机乘客购票后按期搭机的概率为 p，各乘客的行动假定是独立的，试问一架 200 座飞机售出 202 张机票不发生超座的概率。对 $p=0.99$，0.98，0.97，0.96，0.95，计算上述概率.

33. 若母体 ξ 的数学期望 $E\xi = m$，$D\xi = \sigma^2$，抽容量为 n 的子样求其平均值 $\bar{\xi}$，为使 $P\{|\bar{\xi} - m| < 0.1\sigma\} \geqslant 95\%$，问 n 应取多大值？

34. 若 $\{\xi_n, n=1, 2, \cdots\}$ 为相互独立随机变量序列，具有相同分布

$$P\{\xi_n = 1\} = \frac{1}{2}, \quad P\{\xi_n = 0\} = \frac{1}{2}$$

而 $\eta_n = \sum_{k=1}^{n} \frac{\xi_k}{2^k}$，试证 η_n 的分布收敛于 $[0, 1]$ 上的均匀分布.

35. 用特征函数法证明二项分布的泊松逼近定理.

36. 用特征函数法证明，泊松分布当 $\lambda\to\infty$ 时，渐近正态分布.

*37. 若 $\{X_i\}$ 是独立同分布随机变量序列，其分布分别为：(1) $[-a, a]$ 上均匀分布；(2) 泊松分布；(3) Γ 分布，记

$$Y_n=\frac{\sum_{i=1}^{n}(X_i-EX_i)}{\sqrt{\sum_{i=1}^{n}DX_i}}$$

试计算 Y_n 的特征函数，并求 $n\to\infty$ 时的极限.

38. 设 $\{X_n\}$ 独立同分布，$P\{X_n=2^{k-2\ln k}\}=2^{-k}(k=1,2,\cdots)$,则大数定律成立.

*39. 若 $\{X_i\}$ 是相互独立的随机变量序列，均服从 $N(0,1)$,试证

$$W_n=\sqrt{n}\,\frac{X_1+\cdots+X_n}{X_1^2+\cdots+X_n^2}\quad 及\quad U_n=\frac{X_1+\cdots+X_n}{\sqrt{X_1^2+\cdots+X_n^2}}$$

渐近正态分布 $N(0,1)$.

*40. 设 X_1，X_2，…是独立随机变量序列，均服从 $[0, 1]$ 均匀分布，令

$$Z_n=\left(\prod_{i=1}^{n}X_i\right)^{1/n}$$

试证 $Z_n\xrightarrow{P}C$，这里 C 是常数，并求 C.

*41. 若 $\{X_i\}$ 是独立同分布随机变量序列，$EX_i=m$，若 $f(x)$ 是一个有界的连续函数，试证

$$\lim_{n\to\infty}E\left[f\left(\frac{X_1+\cdots+X_n}{n}\right)\right]=f(m)$$

*42. 若 $\{X_i\}$ 是独立同分布、具有有限二阶矩的随机变量序列，试证

$$\frac{2}{n(n+1)}\cdot\sum_{i=1}^{n}iX_i\xrightarrow{P}EX_1$$

*43. 设 X_1，X_2，…相互独立，均服从柯西分布 $p(x)=\frac{1}{\pi}\cdot\frac{1}{1+x^2}$，试证它们不满足格涅坚科关于大数定律的充要条件（见本章习题 9)，即要指出，当 $n\to\infty$ 时

$$E\left[\frac{\left(\sum_{i=1}^{n}X_i\right)^2}{n^2+\left(\sum_{i=1}^{n}X_i\right)^2}\right]\not\to 0$$

*44.（维尔斯特拉斯定理的概率论证明） 设 $f(x)$ 是 $[0,1]$ 上连续函数，利用概率论方法证明：必存在多项式序列 $\{B_n(x)\}$，在 $[0,1]$ 上一致收敛于 $f(x)$.（提示：定义伯恩斯坦多项式

$$B_n(x)=\sum_{m=0}^{n}\binom{n}{m}x^m(1-x)^{n-m}f\left(\frac{m}{n}\right)$$

并利用大数定律.）

45. 设 $\{X_i\}$ 是独立随机变量序列，试证 $X_n\xrightarrow{a.s.}0$ 的充要条件为对任意 $\varepsilon>0$，有 $\sum_{n=1}^{\infty}P\{|X_n|\geqslant\varepsilon\}<\infty$.

46. 试证独立同分布随机变量序列，若存在有限的四阶中心矩，则强大数定律成立.

47. 习题 6 的各个独立随机变量序列是否满足强大数定律？

48. 举例说明博雷尔—康特立引理（i）之逆不成立.

49. 设 $\{X_n\}$ 是相互独立且具有有限方差的随机变量序列，若

$$\sum_{n=1}^{\infty}\frac{DX_n}{n^2}<\infty$$

则必有

$$\lim_{n\to\infty}\frac{1}{n^2}\sum_{k=1}^{n}DX_k=0$$

*50. 设 $f(x)$ 和 $g(x)$ 在闭区间 $[0,1]$ 上连续，且满足 $0\leqslant f(x)<Cg(x)$，这里 C 是一个正常数，则成立

$$\lim_{n\to\infty}\int_0^1\int_0^1\cdots\int_0^1\frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{g(x_1)+g(x_2)+\cdots+g(x_n)}\mathrm{d}x_1\mathrm{d}x_2\cdots\mathrm{d}x_n$$

$$=\frac{\int_0^1 f(x)\mathrm{d}x}{\int_0^1 g(x)\mathrm{d}x}$$

**51. 直接证明科尔莫戈罗夫不等式.

**52. 对事件序列 $\{A_n\}$，及整数序列 $1=n_1<n_2<\cdots$，证明：

（1）$\bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty}A_n=\bigcap_{k=1}^{\infty}\bigcup_{j=k}^{\infty}\bigcup_{n=n_j}^{n_{j+1}-1}A_n$；

（2）若 $\sum_{j=1}^{\infty}P\{\bigcup_{n=n_j}^{n_{j+1}-1}A_n\}<\infty$，则 $P\{\overline{\lim_{n\to\infty}}A_n\}=0$；

（3）对随机变量序列 $\{\zeta_n\}$，若级数

$$\sum_{j=1}^{\infty} P\{\max_{n_j \leqslant n < n_{j+1}} |\zeta_n| \geqslant \varepsilon\}$$

对一切 $\varepsilon > 0$ 收敛，则 $P\{\zeta_n \to 0\} = 1$.

**53. 若 $\{\xi_k\}$ 是独立随机变量序列，方差有限，记

$$S_n = \sum_{k=1}^{\infty} (\xi_k - E\xi_k), \quad \eta_n = \frac{1}{n} S_n$$

(1) 利用科尔莫戈罗夫不等式证明：

$$p_m \equiv P\{\max |\eta_n| \geqslant \varepsilon, \quad 2^m \leqslant n < 2^{m+1}\}$$

$$\leqslant \frac{1}{(2^m \varepsilon)^2} \sum_{j < 2^{m+1}} D\xi_j$$

(2) 对上述 p_m，证明若 $\sum_{k=1}^{\infty} \frac{D\xi_k}{k^2} < \infty$，则 $\sum_{m=1}^{\infty} p_m$ 收敛；

(3) 利用上题结果证明对 $\{\xi_n\}$ 成立科尔莫戈罗夫强大数定律.

**54. (1) 设 $\{c_k\}$ 为常数列，令

$$s_n = \sum_{k=1}^{n} c_k, \quad b_m = \sup\{|s_{m+k} - s_m|, k = 1, 2, \cdots\}$$

$$b = \inf\{b_m, m = 1, 2, \cdots\}$$

试证 $\sum_{k=1}^{\infty} c_k$ 收敛的充要条件是 $b = 0$；

(2) (Kronecker 引理) 对实数列 $\{c_k\}$，若 $\sum_{k=1}^{\infty} \frac{c_k}{k}$ 收敛，则 $\frac{1}{n}\sum_{k=1}^{n} c_k \to 0$.

**55. 若 $\{\xi_k\}$ 是独立随机变量序列，方差有限，记

$$S'_n = \sum_{k=1}^{n} \frac{\xi_k - E\xi_k}{k}$$

(1) 利用科尔莫戈罗夫不等式证明

$$P\{\max_{1 \leqslant k \leqslant n} |S'_{m+k} - S'_m| \geqslant \varepsilon\} \leqslant \frac{1}{\varepsilon^2} \sum_{k=m+1}^{m+n} \frac{D\xi_k}{k^2}$$

(2) 利用上题结论(1)证明：若 $\sum_{k=1}^{\infty} \frac{D\xi_k}{k^2} < \infty$，则 S'_n 以概率 1 收敛.

(3) 利用 Kronecker 引理导出科尔莫戈罗夫强大数定律.

56. 设 $X_1, X_2, \cdots$ 是独立随机变量序列，对它成立中心极限定理，则对 $\{X_n\}$ 成立大数定律的充要条件为 $D(X_1 + \cdots + X_n) = o(n^2)$.

57. 设 $X_1, X_2, \cdots$ 是独立同分布随机变量序列，且 $\dfrac{\sum_{k=1}^{n} X_k}{\sqrt{n}}$ 对每一个 $n = 1, 2,$

…有相同分布，那么，若 $EX_i=0$，$DX_i=1$，则 X_i 必须是 $N(0,1)$ 变量.

58. 设 $\{X_k\}$ 是独立随机变量序列，且 X_k 服从 $N(0,2^{-k})$，试证序列 $\{X_k\}$：(1) 成立中心极限定理；(2) 不满足费勒条件；(3) 不满足林德贝格条件，从而说明林德贝格条件并不是中心极限定理成立的必要条件.

59. 若 $\{X_k\}$ 是独立随机变量序列，X_1 服从 $[-1,1]$ 均匀分布，对 $k=2,3,\cdots$，X_k 服从 $N(0,2^{k-1})$，证明对 $\{X_k\}$ 成立中心极限定理但不满足费勒条件.

*60. 在泊松试验中，第 i 次试验时事件 A 出现的概率为 p_i，不出现的概率为 q_i，各次试验是独立的，以 ν_n 记前 n 次试验中事件 A 出现的次数. 试证：

(1) $\dfrac{\nu_n-E\nu_n}{n}\xrightarrow{P}0$；

(2) 对 $\{\nu_n\}$ 成立中心极限定理的充要条件是 $\sum\limits_{i=1}^{\infty}p_iq_i=+\infty$.

61. 设 $\{X_k\}$ 是独立随机变量序列，X_k 服从 $[-k,k]$ 均匀分布，问对 $\{X_k\}$ 能否用中心极限定理？

62. 试问对下列独立随机变量序列，李雅普诺夫定理是否成立？

(1) $X_k:\begin{pmatrix}-\sqrt{k} & \sqrt{k}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}$；　(2) $X_k:\begin{pmatrix}-k^{a} & 0 & k^{a}\\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}$，$a>0$.

*63. 求证：当 $n\to\infty$ 时

$$\frac{\left(\frac{n}{2}\right)^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)}\int_0^{1+t\sqrt{\frac{2}{n}}}z^{\frac{n}{2}-1}e^{-\frac{nz}{2}}dz\to\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{t}e^{-x^2/2}dx$$

*64. 设 $s>0$，独立随机变量序列 $\{\xi_k\}$，对一切 k，ξ_k 以概率 $\frac{1}{2}$ 分别取值 $\pm k^s$，试利用中心极限定理证明：当 $s\geqslant\frac{1}{2}$ 时，大数定律不成立。(参看习题 4)

*65. 求证：当 $n\to\infty$ 时，$e^{-n}\sum\limits_{k=0}^{n}\dfrac{n^k}{k!}\to\dfrac{1}{2}$.

全书小结

本书从大量重复试验中事件出现频率的稳定性这一经验事实出发，引进了一系列概念，建立了既有广泛实际应用又有深刻理论结果的一整套数学理论；最后这个理论又令人信服地解释了作为出发点的经验事实，到此为止，这个课程已告一段落.

读者可以看到，作为概率论研究对象的随机现象，在自然界和人类社会中是普遍存在的，这既说明了概率论理论的重要性，也决定了它的应用的广泛性.

读者也可以看到，概率论是数学的一个有特色的分支．一方面，由于它与其他数学分支（它们都研究决定性现象）研究对象的不同，因此它有着别开生面的研究课题，从而也有着自己独特的概念与方法．另一方面，它又是一个严谨的数学分支，它的概念有明确的定义，它的方法是严格的，它的结果是深刻的，这是许多著名数学家长期努力的结果，也得益于不断地从其他数学分支吸取有用的概念与方法.

把概率论最基本的概念——事件，概率，随机变量与数学期望分别看作是集合，规范化测度，可测函数与可测函数关于规范化测度的积分，这种观点自从概率论的公理化结构体系出现之后，已被普遍接受．这里强调的是概率论的测度论特点，对于概率论基本概念的明确定义是至关重要的.

读者应当认识到，上述观点的形成是一个历史过程，同时也不应当忘记这些概念的现实背景，这时研究几个经典的模型是很有帮助的．历史上，古典概型、几何概率、伯努利概型等几个模型孕育了早期的概率论，对后来的发展也有重大影响．即使在今天，这些既直观又具体的模型也还值得我们特别注意.

统计独立性是概率论中特有的概念，它的引进大大丰富了概率论的研究．概率论基础中最深入的结果大都是在独立性的假定下获得的，这主要是指几种形式的极限定理：大数定律、强大数定律与中心极限定理．

条件概率是另一重要概念，它使我们能充分利用有关的信息，在概率计算中十分有用．当进一步研究非独立的场合时，条件概率将起更大的作用．

随机变量概念的普遍使用相对来讲还是近代的事，但是这个概念是很自然的，又便于运算，用它来描述事件有不少好处．涉及随机变量的最重要问题中大多数可以通过分布函数来表述，概率论的这一部分可以独立于它的测度论基础而进行，这里所用的主要是分析方法．

数学期望是概率论中最古老的概念，它的明确的直观含意和良好的数学性质使它在概率论中一直占据着重要的地位，各种重要的数字特征大都是某种数学期望，甚至概率也是一种数学期望．数学期望，以及与它相辅相成的方差，刻画了随机变量的概貌，是实际应用中最关心的两个量．

极限定理是概率论中最重要的理论结果，本书通过三个层次介绍了经典结果，相当详尽．分布函数、矩和特征函数这三者是解决古典极限定理的主要工具．

总的说来，本书通过经典模型来提供有关背景，采用公理化结构以明确定义概念，强调独立性以突出学科的特点，利用分析方法来获得深刻的结果，最后也是最重要的一点是，试图通过大量的实例来介绍概率论的日益广泛的应用．

正如本书的书名所指出的，这些知识仅仅是概率论的基础，它们是新的研究的出发点．

参考书目

[1] 格涅坚科. 概率论教程. 丁寿田译. 北京：高等教育出版社，1956

[2] 费勒. 概率论及其应用. 第一卷，胡迪鹤，林向清译. 第二卷，刘文译. 北京：科学出版社，1964，1994

[3] 郑绍濂，吴立德，陶宗英，汪嘉冈. 概率论与数理统计. 第二版. 上海：上海科学技术出版社，1961

[4] 费史. 概率论及数理统计. 王福保译. 上海：上海科学技术出版社，1962

[5] 王梓坤. 概率论基础及其应用. 北京：科学出版社，1976

[6] 钟开莱. 初等概率论附随机过程，魏宗舒等译. 北京：人民教育出版社，1979

[7] 李贤平，卞国瑞，吴立鹏. 概率论与数理统计简明教程. 北京：高等教育出版社，1988

[8] 欣钦. 公用事业理论的数学方法. 张里千，殷涌泉译. 北京：科学出版社，1958

[9] 雅格洛姆，雅格洛姆. 概率与信息. 吴茂森译. 上海：上海科学技术出版社，1964

[10] 伊藤清. 確率論. 東京：岩波書店，1953

[11] 淺野長一郎，江島伸興，李賢平. 基本統計學. 東京：森北出版株式會社，1993

[12] E Parzen. Modern Probability Theory and Its Applications. New York：John Wiley & Sons，1960

[13] J V Uspensky. Introduction to Mathematical Probability. New York：McGraw-Hill，1937

[14] M E Munroe. Theory of Probability. New York：McGraw-Hill，1951

[15] A Rényi. Probability Theory. New York：North-Holland，1970

[16] V K Rohatgi. An Introduction to Probability Theory and Mathematical Statistics. New York：John Wiley & Sons，1976

[17] E Lukacs. Characteristic Functions. London: Charles Griffin, 1960
[18] E Lukacs and R G Laha. Applications of Characteristic Functions. London: Charles Griffin, 1964
[19] R Z Yeh. Modern Probability Theory. New York: Harper & Row, 1973
[20] A Renyi. Wahrscheinlichkeitsrechnung. Berlin: VEB Deutscher Verlag Der Wissenschaften, 1962
[21] В С Пугачев. Теория Случайных Функций и её Применение к Задачам Автоматического Управления. Москва: Физматгиз, 1960

附录一

常用分布表

分布名称	概率分布与密度函数 $p(x)$	数学期望	方差	特征函数	有关章节
退化分布 $I(x-c)$	$p_c=1$ (c 为常数)	c	0	e^{ict}	三§1;四§5;五§2，§3.
伯努利分布（两点分布）	$p_k=\begin{cases}q, k=0\\ p, k=1\end{cases}$ $0<p<1, q=1-p$	p	pq	$pe^{it}+q$	二§3;三§1;四§1,§2;五§1.
二项分布 $B(n, p)$	$b(k;n,p)=\binom{n}{k}p^kq^{n-k}$ $k=0,1,\cdots,n$ $0<p<1, q=1-p$	np	npq	$(pe^{it}+q)^n$	一§3；二§3,§4；三§1；四§1，§2，§4，§5；五§1.
泊松分布 $P(\lambda)$	$p(k;\lambda)=\frac{\lambda^k}{k!}e^{-\lambda}, k=0,1,2,\cdots$ $\lambda>0$	λ	λ	$e^{\lambda(e^{it}-1)}$	二§4;三§1;四§1，§2，§4，§5.
几何分布	$g(k;p)=q^{k-1}p$ $k=1,2,\cdots$ $0<p<1, q=1-p$	$\frac{1}{p}$	$\frac{q}{p^2}$	$\frac{pe^{it}}{1-qe^{it}}$	一§3;二§3;三§1;四§1,§4.

超几何分布	$p_k = \dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$, $M \leqslant N$, $n \leqslant N$, M,N,n 正整数 $k = 0,1,2,\cdots,\min(M,N)$	$\dfrac{nM}{N}$	$\dfrac{nM}{N}\left(1-\dfrac{M}{N}\right)\cdot\dfrac{N-n}{N-1}$	$\sum\limits_{k=0}^{n}\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}e^{itk}$	一§3;三§1;四§1, §2, §4.
巴斯卡分布	$p_k = \binom{k-1}{r-1}p^r q^{k-r}, k = r, r+1, \cdots$ $0 < p < 1, q = 1-p, r$ 正整数	$\dfrac{r}{p}$	$\dfrac{rq}{p^2}$	$\left(\dfrac{pe^{it}}{1-qe^{it}}\right)^r$	二§3; 三§1.
负二项分布	$p_k = \binom{-r}{k}p^r(-q)^k, k = 0,1,2,\cdots$ $0 < p < 1, q = 1-p, r > 0$	$\dfrac{rq}{p}$	$\dfrac{rq}{p^2}$	$\left(\dfrac{p}{1-qe^{it}}\right)^r$	三§1.
正态分布(高斯分布) $N(a, \sigma^2)$	$p(x) = \dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-a)^2}{2\sigma^2}}$ $-\infty < x < \infty, a, \sigma > 0$,常数	a	σ^2	$e^{iat-\frac{1}{2}\sigma^2 t^2}$	三§1,§2,§3;四§1,§2,§3,§5,§6;五§1,§3, §5.
均匀分布 $U[a, b]$	$p(x) = \begin{cases}\dfrac{1}{b-a}, & a \leqslant x \leqslant b\\ 0, & \text{其他}\end{cases}$ $a < b$,常数	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{itb}-e^{ita}}{it(b-a)}$	一§4; 三§1,§2,§3;四§1,§2,§3;五§3.
指数分布	$p(x) = \begin{cases}\lambda e^{-\lambda x}, & x \geqslant 0\\ 0, & x < 0\end{cases}$ $\lambda > 0$,常数	λ^{-1}	λ^{-2}	$\left(1-\dfrac{it}{\lambda}\right)^{-1}$	三§1; 四§1,§3.

续表

分布名称	概率分布与密度函数 $p(x)$	数学期望	方　差	特征函数	有关章节
χ^2 分布	$p(x)=\begin{cases}\dfrac{1}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-x/2}, & x\geqslant 0\\ 0, & x<0\end{cases}$ n 正整数	n	$2n$	$(1-2it)^{-n/2}$	三§3.
Γ 分布 $G(\lambda, r)$	$p(x)=\begin{cases}\dfrac{\lambda^r}{\Gamma(r)}x^{r-1}e^{-\lambda x}, & x\geqslant 0\\ 0, & x<0\end{cases}$ $r>0, \lambda>0$ 常数	$r\lambda^{-1}$	$r\lambda^{-2}$	$\left(1-\dfrac{it}{\lambda}\right)^{-r}$	三§1，§3. 四§5.
柯西分布	$p(x)=\dfrac{1}{\pi}\cdot\dfrac{\lambda}{\lambda^2+(x-\mu)^2}$ $-\infty<x<\infty, \lambda>0, \mu$ 常数	不存在	不存在	$e^{i\mu t-\lambda\|t\|}$	三§3; 四§1.
t 分布	$p(x)=\dfrac{\Gamma\left(\dfrac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\dfrac{n}{2}\right)}\left(1+\dfrac{x^2}{n}\right)^{-(n+1)/2}$ $-\infty<x<\infty$ n 正整数	$0(n>1)$	$\dfrac{n}{n-2}(n>2)$		习题三

帕累托分布	$p(x)=\begin{cases} rA^r\dfrac{1}{x^{r+1}}, & x\geqslant A \\ 0, & x<A \end{cases}$ $r>0, A>0$	($r>1$ 时存在)	($r>2$ 时存在)		习题四
F 分布	$p(x)=\begin{cases} \dfrac{\Gamma\left(\dfrac{k_1+k_2}{2}\right)}{\Gamma\left(\dfrac{k_1}{2}\right)\Gamma\left(\dfrac{k_2}{2}\right)}k_1^{k_1/2}k_2^{k_2/2}\cdot \\ \dfrac{x^{k_1/2-1}}{(k_2+k_1x)^{(k_1+k_2)/2}}, & x\geqslant 0 \\ 0, & x<0 \end{cases}$ k_1,k_2 正整数	$\dfrac{k_2}{k_2-2}$ $(k_2>2)$	$\dfrac{2k_2^2(k_1+k_2-2)}{k_1(k_2-2)^2(k_2-4)}$ $(k_2>4)$		习题三
β 分布	$p(x)=\begin{cases} \dfrac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)}x^{p-1}(1-x)^{q-1}, \\ \qquad 0<x<1 \\ 0, \qquad x\leqslant 0 \text{ 或 } x\geqslant 1 \end{cases}$ $p>0, q>0$ 常数	$\dfrac{p}{p+q}$	$\dfrac{pq}{(p+q)^2(p+q+1)}$	$\dfrac{\Gamma(p+q)}{\Gamma(p)}\cdot\sum\limits_{j=0}^{\infty}\dfrac{\Gamma(p+j)(\mathrm{i}t)^j}{\Gamma(p+q+j)\Gamma(j+1)}$	三 § 3
对数正态分布	$p(x)=\begin{cases} \dfrac{1}{\sigma x\sqrt{2\pi}}\mathrm{e}^{-\frac{(\ln x-a)^2}{2\sigma^2}}, & x>0 \\ 0, & x\leqslant 0 \end{cases}$ $a,\sigma>0$ 常数	$\mathrm{e}^{a+\sigma^2/2}$	$\mathrm{e}^{2a+\sigma^2}(\mathrm{e}^{\sigma^2}-1)$		习题三
威布尔分布	$p(x)=\begin{cases} \alpha\lambda x^{\alpha-1}\mathrm{e}^{-\lambda x^{\alpha}}, & x>0 \\ 0, & x\leqslant 0 \end{cases}$ $\lambda>0, \alpha>0$ 常数	$\Gamma\left(\dfrac{1}{\alpha}+1\right)\lambda^{-1/\alpha}$	$\lambda^{-2/\alpha}\left[\Gamma\left(\dfrac{2}{\alpha}+1\right)-\left(\Gamma\left(\dfrac{1}{\alpha}+1\right)\right)^2\right]$		习题三

附 录 二

泊松分布 $P\{\xi=r\}=\frac{\lambda^r}{r!}e^{-\lambda}$的数值表

r	λ							
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
0	0.904837	0.818731	0.740818	0.670320	0.606531	0.548812	0.496585	0.449329
1	0.090484	0.163746	0.222245	0.268128	0.303265	0.329287	0.347610	0.359463
2	0.004524	0.016375	0.033337	0.053626	0.075816	0.098786	0.121663	0.143785
3	0.000151	0.001092	0.003334	0.007150	0.012636	0.019757	0.028388	0.038343
4	0.000004	0.000055	0.000250	0.000715	0.001580	0.002964	0.004968	0.007669
5	—	0.000002	0.000015	0.000057	0.000158	0.000356	0.000696	0.001227
6	—	—	0.000001	0.000004	0.000013	0.000036	0.000081	0.000164
7	—	—	—	—	0.000001	0.000003	0.000008	0.000019
8	—	—	—	—	—	—	0.000001	0.000002

续表

r	λ							
	0.9	1.0	1.5	2.0	2.5	3.0	3.5	4.0
0	0.406570	0.367879	0.223130	0.135335	0.082085	0.049787	0.030197	0.018316
1	0.365913	0.367879	0.334695	0.270671	0.205212	0.149361	0.150091	0.073263
2	0.164661	0.183940	0.251021	0.270671	0.256516	0.224042	0.184959	0.146525
3	0.049398	0.061313	0.125510	0.180447	0.213763	0.224042	0.215785	0.195367
4	0.011115	0.015328	0.047067	0.090224	0.133602	0.168031	0.188812	0.195367
5	0.002001	0.003066	0.014120	0.036089	0.066801	0.100819	0.132169	0.156293
6	0.000300	0.000511	0.003530	0.012030	0.027834	0.050409	0.077098	0.104196
7	0.000039	0.000073	0.000756	0.003437	0.009941	0.021604	0.038549	0.059540
8	0.000004	0.000009	0.000142	0.000859	0.003106	0.008102	0.016865	0.029770
9	—	0.000001	0.000024	0.000191	0.000863	0.002701	0.006559	0.013231
10	—	—	0.000004	0.000038	0.000216	0.000810	0.002296	0.005292
11	—	—	—	0.000007	0.000049	0.000221	0.000730	0.001925
12	—	—		0.000001	0.000010	0.000055	0.000213	0.000642
13	—	—	—	—	0.000002	0.000013	0.000057	0.000197
14	—	—	—	—	—	0.000003	0.000014	0.000056
15	—	—	—	—	—	0.000001	0.000003	0.000015
16	—	—	—	—	—	—	0.000001	0.000004
17	—	—	—	—	—	—	—	0.000001

续表

r	λ						
	4.5	5.0	6.0	7.0	8.0	9.0	10.0
0	0.011109	0.006738	0.002479	0.000912	0.000335	0.000123	0.000045
1	0.049990	0.033690	0.014873	0.006383	0.002684	0.001111	0.000454
2	0.112479	0.084224	0.044618	0.022341	0.010735	0.004998	0.002270
3	0.168718	0.140374	0.089235	0.052129	0.028626	0.014994	0.007567
4	0.189808	0.175467	0.133853	0.091226	0.057252	0.033737	0.018917
5	0.170827	0.175467	0.160623	0.127717	0.091604	0.060727	0.037833
6	0.128120	0.146223	0.160623	0.149003	0.122138	0.091090	0.063055
7	0.082363	0.104445	0.137677	0.149003	0.139587	0.117116	0.090079
8	0.046329	0.065278	0.103258	0.130377	0.139587	0.131756	0.112599
9	0.023165	0.036266	0.068838	0.101405	0.124077	0.131756	0.125110
10	0.010424	0.018133	0.041303	0.070983	0.099262	0.118580	0.125110
11	0.004264	0.008242	0.022529	0.045171	0.072190	0.097020	0.113736
12	0.001599	0.003434	0.011264	0.026350	0.048127	0.072765	0.094780
13	0.000554	0.001321	0.005199	0.014188	0.029616	0.050376	0.072908
14	0.000178	0.000472	0.002288	0.007094	0.016924	0.032384	0.052077
15	0.000053	0.000157	0.000891	0.003311	0.009026	0.019431	0.034718
16	0.000015	0.000049	0.000334	0.001448	0.004513	0.010930	0.021699
17	0.000004	0.000014	0.000118	0.000596	0.002124	0.005786	0.012764
18	0.000001	0.000004	0.000039	0.000232	0.000944	0.002893	0.007091
19	—	0.000001	0.000012	0.000085	0.000397	0.001370	0.003732
20	—	—	0.000004	0.000030	0.000159	0.000617	0.001866
21	—	—	0.000001	0.000010	0.000061	0.000264	0.000889
22	—	—	—	0.000003	0.000022	0.000108	0.000404

正态分布密度函数 $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ 及

分布函数 $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt$ 的数值表

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
0.0	0.398942	0.500000	2.3	0.028327	0.989276
0.1	0.396952	0.539828	2.4	0.022395	0.991802
0.2	0.391043	0.579260	2.5	0.017528	0.993790
0.3	0.381388	0.617911	2.6	0.013583	0.995339
0.4	0.368270	0.655422	2.7	0.010421	0.996533
0.5	0.352065	0.691462	2.8	0.007915	0.997445
0.6	0.333225	0.725747	2.9	0.005953	0.998134
0.7	0.312254	0.758036	3.0	0.004432	0.998650
0.8	0.289692	0.788145	3.1	0.003267	0.999032
0.9	0.266085	0.815940	3.2	0.002384	0.999313
1.0	0.241971	0.841345	3.3	0.001723	0.999517
1.1	0.217852	0.864334	3.4	0.001232	0.999663
1.2	0.194186	0.884930	3.5	0.000873	0.999767
1.3	0.171369	0.903200	3.6	0.000612	0.999841
1.4	0.149727	0.919243	3.7	0.000425	0.999892
1.5	0.129518	0.933193	3.8	0.000292	0.999928
1.6	0.110921	0.945201	3.9	0.000199	0.999952
1.7	0.094049	0.955435	4.0	0.000134	0.999968
1.8	0.078950	0.964070	4.1	0.000089	0.999979
1.9	0.065616	0.971283	4.2	0.000059	0.999987
2.0	0.053991	0.977250	4.3	0.000039	0.999991
2.1	0.043984	0.982136	4.4	0.000025	0.999995
2.2	0.035475	0.986097	4.5	0.000016	0.999997

[General Information]
[illegible]=[illegible]
[illegible]=[illegible]
[illegible]=339
SS[illegible]=10070152
[illegible]=1997[illegible]04[illegible]2[illegible]

[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible] [illegible]
1. [illegible]
2. [illegible]
3. [illegible]
4. [illegible]
5. [illegible]
[illegible]
[illegible]
[illegible] [illegible]
1. [illegible]
2. [illegible]
3. [illegible]
4. [illegible]
[illegible]
[illegible]
[illegible] [illegible]
1. [illegible]
2. [illegible]
3. [illegible]
[illegible]
[illegible]
[illegible] [illegible]
1. [illegible]
2. [illegible]
3. [illegible]
4. [illegible]
5. [illegible]
6. [illegible]
[illegible]
[illegible]
[illegible] [illegible]
1. [illegible]
2. [illegible]
3. [illegible]
4. [illegible]
5. [illegible]
[illegible]
[illegible]

□□□□
□□□□
□□□ □□□□□
□□□
□□□□ P{ξ□r}□λr/r!e-λ □□□□
□□□□□□□□φ(x)=□□□□□□ Φ(x)□□□□□
□□□